

Two successful weeks at Rio

Rumours that the UN conference that finished last Sunday in Rio was a failure are false. The urgent matters were dealt with adequately, while the global character of other problems has been more widely acknowledged.

By the modest tests this journal has been urging on its readers for the past several months, the UN Conference on Environment and Development that ended last Sunday was a huge success. The proof is that even the United States has signed the global warming treaty, providing a framework within which greenhouse gases can be progressively regulated and restricted. That is important. Almost as many governments (but not the United States) signed a treaty to protect biodiversity so watered down (as the enthusiasts complain) that it will do no harm, but will serve both as a proof of an international readiness to tackle problems that arise and an incentive to interested parties to put energy (especially intellectual energy) into their accurate definition. On both counts, there has been progress.

So why should there be such an army of people straggling back from Rio de Janeiro loudly complaining that nothing worthwhile was done there? Most simply, they expected too much — the eradication of poverty from the poor countries, the voluntary abdication of wealth elsewhere, an understanding that poor countries may appoint themselves game-wardens to the world and then charge back what they consider to be the costs they have incurred and, of course, specific limits on greenhouse gases, however inappropriate or unenforceable. And why have so many well-intentioned people been misled? In part, the resonance of their utopian rhetoric is self-amplifying, but they have also been misled by others. Mr Maurice Strong, the Canadian secretary of the UN conference (the Ross Perot of the environmental movement?), has a particularly heavy burden of the responsibility for that state of affairs.

Framework

What is to happen now? Constitutionally, the two treaties have been signed, will become binding on signatories when 50 states have ratified them and will then take on a life of their own, with regular meetings of governments that are party to them. In each case, there will be annual meetings of the signatory governments, at each of which it will in principle be possible for the signatories to adopt protocols that put stuffing into the framework treaties. That is the procedure followed under the Vienna Treaty on the Protection of the Ozone Layer. That it can function well is demonstrated by the existence of the Montreal Protocol. But everything will hang on the advice the parties to the new treaties are given.

The global warming treaty, for example, will eventually have a permanent secretariat of its own, as well as two

permanent technical bodies responsible respectively for data and understanding on climatic change and for monitoring and assessing various countries' plans for abating emissions. That these bodies should be competent goes without saying. But they should also be intelligent and imaginative — much more so than the Intergovernmental Panel on Climate Change (IPCC). The danger with the global warming treaty is that it could be quickly turned into a drill for assigning allowable quotas for emissions to its members, forgetting that the real task is that of abating the greenhouse effect at the least cost. So how will it be ensured that the avoidance of a tonne of methane emission earns as much credit as the avoidance of (roughly) 15,000 tonnes of carbon dioxide? Or that a European government might be allowed to clean up less efficient plants elsewhere rather than its own relatively efficient fuel-burning equipment?

Biodiversity

The biodiversity treaty has a harder path to travel, if only because its objectives are so muddled. That does not mean that diversity is a bad thing — far from it — but that its value must be more carefully defined, as a function of place and character, than it has been so far. There are some ecosystems, the Amazon Basin for example, which by their scale and distinctiveness are unique on the surface of the Earth (and which may, if destroyed, have implications for global warming). What is their value, not in numbers, but in quality?

As a source of medicines not yet dreamed of? As a case study of the wonders of evolution by natural selection since Gondwanaland? As a habitat for primitive people, apparently illiterate and not especially healthy, in a kind of living anthropological museum? And granted that Brazilians' slash-and-burn techniques for settling the Amazon are both needlessly destructive and short-sighted, what ways, if any, of exploiting the region are compatible with its conservation? And how much would the rest of the world pay in perpetual compensation to see that done? Or does the Amazon Basin belong not to Brazil, but to its aboriginal Indians, and what then should be done for them (and who will tell the government of Brazil?)?

Such questions have not been thought through carefully enough to be asked confidently. They are not only technical in character, but political, economic and legal as well. As things are, they are probably unanswerable. The best bet for the Amazon Basin is a moratorium on development of the kind that might be bought by the imaginative conversion of



some of Brazil's overseas debt into obligations of restraint. Countries such as the Netherlands have already embarked along that route. The biodiversity treaty is a framework within which more comprehensive agreements might be reached, for the Amazon and elsewhere. But everything is a special case, requiring special study. That should be the cornerstone of the way in which the biodiversity treaty functions. Meanwhile, Mr John Major, the British prime minister, is surely right to advocate a programme of more vigorous taxonomic research in regions such as the Amazon (and it is needlessly mean, even by its own standards, for British Friends of the Earth to describe his Rio speech as "empty waffle").

None of this touches the poverty of the poor. Strong's fault is that he has encouraged, especially among the governments of the developing countries, the belief that compensation for custodianship will meet the capital costs of development. That is a gigantic and cruel mistake, especially with a Rio agenda innocent of the issue of population growth. But is not rapid population growth in the developing countries itself a consequence of their poverty? Nobody disputes that in the rich world, the benign demographic transition from high to low rates of birth and death has invariably followed rising prosperity and improved public health. Yet many governments of developing (and quickly growing) countries could be trying harder even as things are. Sooner rather than later, there will have to be a UN conference on that issue as well. □

Moratorium ending

The impending Washington summit has a daunting list of nuclear issues that must be tackled.

Now that the Cold War has ended, why does nuclear testing continue? That is an issue raised this week by the executive committee of the Pugwash Organization which, unlike other international pressure groups, is almost laconic in what it says in public. Specifically, Pugwash has taken fright that it will soon be a year since President Mikhail Gorbachev (remember him?) volunteered, on behalf of the Commonwealth of Independent States, a one-year moratorium on testing. Since then, President Boris Yeltsin had said that Russian tests will resume when the moratorium expires in September. Pugwash asks that the moratorium should be extended at the Washington summit this week, and that the United States should join in.

That, of course, would be an excellent development; there is more than an element of the bizarre in the continued and repeated testing of weapons whose purposes have been confused (or even made nugatory) by the events of the past few months. But, sadly, the summit planned for Washington has even more urgent nuclear business to attend to. Plans for a further bilaterally agreed reduction of strategic arms appear to have foundered on the issue of missiles carrying several warheads, while confusion persists about the role of Russia as the nuclear custodian of other ex-Soviet republics,

the Ukraine conspicuously, but also Khazakstan. Will they eventually become members of the Nuclear Non-Proliferation Treaty (NPT), and if they do, will it be as nuclear or nonnuclear powers?

That question cannot be left unanswered for much longer. Three years from now, the NPT will lapse unless its members elect for its continuation (and for the restrictions it imposes on them). Already there are worrying signs of renewed hankering after independent facilities for making bombs — Iraq last year, reports of attempts illicitly to sell ex-Soviet fissile material by Vienna-based agents only last week. The big danger is that the nonproliferation regime will turn into a leaky sieve long before its sponsors (Russia, presumably, still among them) have worked out a way of selling its virtues to the nonnuclear members of the treaty. A moratorium on testing would help powerfully in that direction. □

Genome propaganda

Concealing the truth without lying is an old art, now spreading in the US human genome project.

By now there can hardly be a researcher who has not heard some account of the unfortunate circumstances behind the resignation in April of Dr James Watson as director of the US human genome project. But the account of the resignation in the current issue of *Human Genome News*, the project's official newsletter, recalls the old propaganda technique of reporting the facts without telling the truth. Nowhere does this account mention what most readers already know from other sources, that Watson resigned after Bernadine Healy, the director of the National Institutes of Health, launched an investigation into his financial holdings. Not even reading between the lines provides a hint of friction between the two.

Instead, the newsletter treats Watson's departure as a routine transition. It quotes him to the effect that directing both the genome project (based in Bethesda, Maryland) and the Cold Spring Harbor Laboratory (in New York state) had become too burdensome for him and his family. It also notes that Watson had told his advisory committee as early as last January — before the controversy arose — that he was thinking of leaving. Although that is true, it is also true that the very press officer who wrote the newsletter account told reporters at the time not to take the comment too seriously; Watson often threatens to resign, she explained.

No one expects the genome project's official newsletter to wallow in gossip. But by pretending that there was no dispute at all — when even Healy was willing to discuss the situation openly — the newsletters belittles its researcher-readers, who are grown-up people and who deserve an accurate and balanced report, and reflects badly on the credibility of the enterprise. Perhaps the lesson from this shabby episode is that the \$7 million a year the United States is spending on the ethics of genome research is not enough. □

European Patent Office rejects bid to revoke first plant patent

Munich. The European Patent Office (EPO) last week rejected an appeal by industrialists, environmentalists and farmers against its decision three years ago to grant the first European plant patent.

The patent, held by the US-based Lubrizol Corporation, covers any transgenic plant originally derived from a plant cell that has been transfected with a plant gene under the control of a plant promoter, using the *agrobacterium* T-DNA as the gene-transfer system. An appeals board within the EPO rejected claims that the technical advance made by Lubrizol was insufficiently novel, and said that it has never been a generally accepted principle in Europe that patents may not be granted for plants. It also dismissed a range of moral objections which, it said, are outside its authority.

Opponents of the Lubrizol patent argue that patent laws that worked happily when the biological arena was confined to microbiological products do not fully come to terms with the complex possibilities offered by biotechnology. They also claim that the EPO's decision gives an unfair market advantage to the patent owner, whose innovative contribution they allege to be minimal. In addition, political interest groups are challenging the basic assumption that the EPO has the right to issue patents on any life-form.

Part of the patent, which claimed rights to the gene-transfer methodology itself, was cancelled because the appeal board agreed that the technology involved was of a general nature. Nevertheless, Lubrizol officials are pleased with the ruling and believe that it strengthens their application for a US patent. In confirming the patentability of plants, Lubrizol says, the EPO has addressed an issue important to all plant biotechnology companies.

All of the interest groups intend to appeal further against the patent. Unlike the US system, where the issued patent is final, European laws allow appeals to be heard after every decision in the process, provided that objections are lodged within nine months.

Further appeals from biotechnology companies, which are themselves applying for patents for transgenic plants, will challenge the modified Lubrizol patent on the grounds that it is based on a technique that they do not believe is innovative. Although it rejected the gene-transfer methodology itself as insufficiently novel to warrant a patent, the EPO ruled that the detection of the expressed gene is suitably novel to warrant patent coverage of all plant cell progeny. But opponents believe that this is simply an obvious corollary of the transfection process.

"We maintain that expression is not enough to support a claim of inventiveness", says Jan van Rompaey, patent manager at Belgium's Plant Genetic Systems and an adviser to the opposition groups. "Having shown that transfection can be achieved, it is fairly obvious that there would be a reasonable chance of transcription of the gene and detection of the product. This is not innovation."

The breadth of the coverage of the patent is also criticized by biotechnology companies. They claim that the rewards to Lubrizol are out of proportion to its scientific input, and that the patent unfairly restricts their own "freedom to operate". Edward Veltkamp, head of research and development for Sandoz Seeds in Basel, Switzerland, says that "a reasonable relationship should exist between the scope of the claims granted and the innovatory contribution delivered by the applicant." But EPO believes that there is no legal relationship between the two.

Environmentalists and other political interest groups are also unhappy about a recent reinterpretation by EPO of an article in the European Patent Convention (EPC),

the rules that the EPO accepts as its guide, which excludes the patenting of any natural life-form. They say that its reinterpretation — which paved the way for Europe's first animal patent, the 'oncomouse' developed by Philip Leder of Harvard University (see *Nature* 353, 589 1991) — is an undemocratic decision affecting fundamental moral and environmental issues. They claim that the EPO does not have the right to make such a decision.

In the Lubrizol patent application, claims were acceptable to the EPO because they referred to plants and not the plant varieties specified in the EPC article. According to Harald Wosihnoj of Austria's Gen-ethisches Netzwerk, a collection of individuals and advocacy groups campaigning for the ethical use of genetic technology, these linguistic gymnastics undermine the EPC's aims.

As debates about life-form patenting gather momentum, opponents agree that the patent laws need to be refined to accommodate the issues raised by new biological technologies. They concede that the EPO has a difficult job in trying to apply rules formulated in the 1960s, before the advent of transgenic life-forms. **Allison Abbott**

US to seek gene patents in Europe

Washington. The US National Institutes of Health (NIH) are this week expected to submit to the European Patent Office a controversial patent application on more than 2,000 cDNA fragments. Coming a year after NIH launched an international dispute with its first patent application for such gene sequences, the European filing is an "interim policy" that will keep NIH's options open while officials debate the alternatives, says Bernadine Healy, the NIH director.

Despite the protests of France, Britain and most of the scientific community, which opposes patents on gene sequences when a biological function is not known, the US government has not yet taken an official position. European patent law allows inventors to file for up to 12 months after a US filing without risk of losing the patent because of public disclosure. US officials appear receptive to the idea of an international treaty to resolve the issue, but no one has suggested what such a treaty should say or how it should be implemented.

Interviewed in Paris last month, Hubert Curien, the French science minister, said that the European Patent Office had assured him that such sequences were not patent-

able. "I will be very sorry if this [attempting to patent in Europe] is what the United States decides", he said at the time.

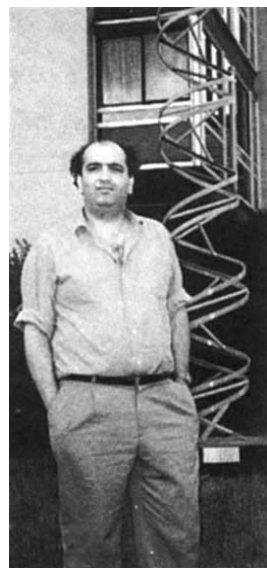
Although French scientists have encouraged President François Mitterrand to raise the issue with the US president, George Bush, Curien said that he had already spoken to White House science adviser D. Allan Bromley, who defended patenting as a way to ensure publication. "I don't think that Mitterrand will get a different answer", Curien said.

In a related development, J. Craig Venter, the NIH scientist who sequenced the fragments that are the subject of the US patent application, was one of three dozen scientists who signed a resolution calling for an end to patents of "naturally occurring gene sequences" in favour of patents only on the uses of those sequences. Venter has said that he supported the NIH patent filing as a way to stimulate debate, but that he hopes the patent will not be granted. The statement, approved by participants in a South-North human genome conference held last month in Brazil, also calls for an international treaty. But it offers no details.

Christopher Anderson

New French genome centre aims to prove that bigger really is better

Paris. Were it not for the big steel double helix outside, the world's largest gene-mapping laboratory might be mistaken for just another office building in one of the industrial parks that dot the countryside south of Paris. But the illusion is quickly dispelled. Above the ground floor, where technicians



Daniel Cohen is winning converts to industrial-scale gene mapping.

teach children with genetic diseases to use voice-operated computers, lie rooms of robots — the makings of an industrial-sized 'experiment in scale' that may be the closest thing to big science that biology has ever seen. In full operation for only a few months, the Généthon, as it is known, has caused genome researchers around the world to re-

think their assumptions about the best way to complete the ambitious project to map and sequence the human DNA molecule.

Last month, Généthon researchers took

the genome world by surprise when they revealed they had already finished a map covering some 28 per cent of the genome and almost the entire chromosome 21. Daniel Cohen, director of the Paris-based Centre d'Etude du Polymorphisme Humain (CEPH) and the driving force behind the Généthon, told a meeting at Cold Spring Harbor Laboratory in New York state that he expects to be able to extend the map to cover 90 per cent of the genome by the end of the year.

But what really caused a stir was the way in which Cohen had collected the data: instead of using scientists, he had technicians running an array of automated gene machines that exist only as prototypes. The Généthon demonstrates an almost blue-collar approach to biology. More assembly line than ivory tower, the laboratory exists mostly to produce genetic information that other scientists will use.

The news that an industrial-scale gene mapping facility was not only up and running but also pumping out data faster than virtually any other group has caused US researchers to question their approach.

"It's impressive", says Mark Guyer, assistant director for program coordination at the National Center for Human Genome Research within the US National Institutes of Health (NIH). "It also confirms what we've been hearing from some of our own people about the advantages of scale. I'm not sure we can set up something that exactly duplicates their set-up, but in terms of learning the lesson of the industrial approach, we can definitely apply those sort

of techniques to US operations."

Another convert is Eric Lander, a mathematician/geneticist and director of the genome centre at the Massachusetts Institute of Technology (MIT)'s Whitehead Institute. The Généthon "has demonstrated the tremendous advantages of scale", he says. Cohen has "done a lot to change people's perceptions. It's no longer unthinkable to take on whole genomic targets" rather than specific disease genes or chromosomes.

Lander, in fact, is more than an admirer; his laboratory has begun to collaborate with Cohen's group. And Cohen is expected to take a part-time position at MIT later this year as the first step in a major collaborative centre (see story below).

The Généthon itself is every bit as startling as its results. The \$10-million central facility is a large room, ringed with 20 robots, surrounding a central climate-controlled computer installation. The robots, which automate the DNA fingerprinting process, were developed by the French company Bertin as part of the European industrial collaboration Eureka Labimap 2000.

Each robot can perform Southern blots — a technique to distinguish DNA fragments by the characteristic bands they leave after gel electrophoresis — on 16 gels (with 20 lanes each) a day. Together, the machines have the potential to process more than 6,000 DNA samples a day. Yet the entire operation requires only five technicians.

Elsewhere in the laboratory, 12 automated sequencers from Applied Biosystems Inc. are producing some 500 cDNA

Will MIT be home to next big centre?

Washington. Riding on the success of the Généthon and its industrial approach to genome mapping, Daniel Cohen wants to set up a companion laboratory at the Massachusetts Institute of Technology (MIT). Although the size of the centre is unclear, negotiations are under way for Cohen to take a part-time appointment at MIT; he has already begun to collaborate with Eric Lander, a geneticist at MIT's Whitehead Institute.

Cohen, director of the Centre d'Etude du Polymorphisme Humain, believes that he and Lander, with \$20 million and 100 to 200 technicians and researchers, could create a US analogue to the Paris-based centre. "When you want to do really well, you get the right environment. The United States is the best for technology development", Cohen says. "We have done the most that is possible in France."

Lander is equally enthusiastic about the collaboration but more cautious about its scope. "I don't know where I'd put another 20 people, much less 200", he says. He confirms that "we're considering a more extensive collaboration", but says that "we'll have a better idea of what we're doing in two months".

The two scientists have much in common. Lander is a

mathematician-turned-geneticist; Cohen got his first degree as an engineer. Both researchers favour what they call a 'whole genome' approach to genetic mapping.

"Clearly chromosomes are important things, but mapping them *per se* requires chromosome-specific reagents, which have biases", Lander explains. "On a pure efficiency basis, whole genome approaches are the best — you can take whole DNA libraries. The reason [Cohen] and I get along so well is that we both agree that the whole-genome approach is the right way to go." He is also taken by the challenge of applying large-scale techniques to the world of small science.

Mark Guyer, assistant director for programme coordination at the US National Center for Human Genome Research, says that the US project has become "philosophically receptive" to several larger centres, such as the one Cohen and Lander have in mind. A year or two ago, such an idea would have outraged biologists concerned about an erosion of support for smaller laboratories. But Lander and Cohen have proved that production-scale approaches can be remarkably efficient, and it is hard to argue with success.

C.A.

sequences a week. A prototype 'spotting machine' (or, more formally, a multi-purpose clone-handling machine) can prepare 300,000 Yeast Artificial Chromosomes (YACs) in a day virtually on its own. And a prototype 'pooling machine' (known around the laboratory as Barbara) uses a special algorithm to combine automatically DNA samples from 12,288 wells into a much smaller number of pooled samples, which are then screened. This pooling process can reduce screening time by a factor of 10.

Surrounded by such technology, Cohen is confident that he can handle the entire genome. He intends to clone random DNA fragments, each more than one million bases long, in the special 'megaYACs' that researchers at CEPH have developed. Combined, this library will cover 90 per cent of the genome.

Généthon cannot do everything — the genome project also requires researchers to fill in the gaps, order the sequences to form a usable map, find genes and their functions and actually read the sequences base by base. But Cohen's accomplishments are impressive, and the megaYACs and library he promises have won applause on both sides of the Atlantic.

Officials from several national programmes (including the United States) see the Généthon as a model for similar facilities in their own countries. The laboratory's industrial approach — a central high-technology mapping resource that can be run by a network of researchers from smaller groups — appears to be a powerful idea. But Cohen warns that the factors that led to the creation of the Généthon may not be easy to reproduce.

Cohen says that the laboratory owes its existence to "four miracles". The first is CEPH, started in 1983 by immunologist and Nobel prizewinner Jean Dausset and Cohen, and its early embrace of industrial approaches. The second was Bertin's willingness to pay for the development costs of robots based on old technology — Southern blotting — rather than waiting for the technology to improve. As it turned out, older technologies, when automated and perfected, can often outperform newer methods.

Perhaps the most important miracle — and the hardest to reproduce elsewhere — was the support of the French muscular dystrophy charity, AFM. Headed by Bernard Barataud, a bearish former electrical engi-



Généthon's robots promise hands-off biology.

Industrial in more than name

Paris. The Généthon may soon have real industrial partners to go with its industrial approach to the genome project. Officials at the massive gene-mapping laboratory are planning to make it the core of a public-private consortium they call a Genepôle. Companies are being invited to set up complementary laboratories near the Généthon in Evry, south of Paris, where industrial researchers would work side by side with Généthon scientists to turn gene-mapping technology into products.

The first such collaboration is expected to be in software for sequence analysis; discussions are already under way with IBM. Another technology nearing commercialization is biological robotics, an area in which French industry has already shown an interest. Daniel Cohen, director of the Centre d'Etude du Polymorphisme Humain (CEPH), which runs the Généthon's mapping effort, says that he hopes to have companies on the site within a year or two. He says that the government would prefer that French companies are involved at the beginning, but so far there seems to be no serious opposition to participation by US companies as well. Most of the interest, in fact, has been from US companies and venture capitalists. Cohen expects that the nonprofit side of the collaboration — CEPH, the French muscular dystrophy charity AFM and the government — would pay for the basic research, and that the companies would pay for the development of marketable technology.

C.A.

neer, AFM has been conducting telethons that typically raise more than \$15 million a year for genetics research, of which about \$13 million goes to the Généthon. The fourth miracle, Cohen says, is that the French genetics community did not try to stop him. His success, he believes, lies in avoiding the kind of politics that would probably sink such an audacious effort in the United States or Britain.

For those thinking of starting their own Généthon, Cohen warns that "finding one miracle is hard enough; another four miracles may be impossible." Nevertheless, the lessons from the project may have broad applications.

Piotr Slonimski, the new director of the French genome project (and one of the principal organizers of the 147-team effort that completed the sequence of the yeast genome last month [see *Nature* 357, 38; 1992]), compares the Généthon model to a mainframe computer with many users. In contrast, the yeast project more closely resembled a network of smaller computers linked together. The secret, he says, is applying the right model to the problem at hand. Although biological analysis is often best done by a network, Slonimski says, physical mapping is better accomplished at central facilities such as the Généthon.

Having proved its thesis that an industrial-scale approach can work in genetics, the Généthon is fast becoming a valuable international resource. Some 20 per cent of its effort is now spent on work for others. More than 50 outside laboratories, including about 20 in France, 30 in the rest of Europe and a few in the United States, now send their samples to the laboratory for analysis. Data sharing does not seem to be a problem; Cohen plans to make his megaYAC libraries available this summer, and he promises not to patent gene sequences in the absence of knowledge of their function.

Despite the acclaim, Cohen is not surprised by the project's success. "This is no-risk research", he says. "This is easy. You put in one dollar and you get one thing; you put in two dollars, you get two things." He says he never doubted that it would work, nor that it could grow.

Although nearly everyone agrees that the laboratory has been an eye-opening achievement, they point out that much of the productivity is still only potential. Getting the long sequences is one thing, ordering them is quite another. "What they have is a plan to make a physical map, but it isn't there yet", says one US researcher.

Some other researchers are simply offended by the whole industrial approach, or, as Cohen puts it, the "changing scale of biology", a complaint he quickly dismisses. "They, too, will have to use robots [some day], and they just hate it", he says. "They will cry for five or ten years and then they will forget it."

Christopher Anderson

Serbs oppose leader

Munich. The Serbian Academy of Science and Arts has called for the resignation of the Serbian president, Slobodan Milosevic, who, they say, makes the plight of Serbia "even graver" in his attempts to put Serbia back on the political map. The appeal is signed by more than 800 scientists and 3,000 other citizens. Belgrade's Institute of Physics, one of the largest in Yugoslavia, has also appealed to the United Nations to ensure that dangerous chemical plants are kept out of range of artillery fire. Its immediate concern is the SODA SA plant in Tuzla, in northeast Bosnia, which stores a hundred times more chlorine than was held at the Union Carbide plant in India at the time of the Bhopal disaster.

Allison Abbott

Computer network rumours prove hard to kill

Washington. It is the rumour that will not die. Every six months or so, subscribers to Internet are warned about a plan by the US Federal Communications Commission (FCC) to impose a tax or other surcharge on modems and telecommunications users. Prompted by the messages to complain about the fee, hundreds of other Internet users fire off angry letters to Washington that describe the dire consequences of such a price increase.

It is an impressive display of citizenship. It is also completely wrong. There is no such proposal for a modem tax or 'usage fee'. There has not been one for half a decade.

"Please help us", begs Kathie Kneff, chief of the FCC's informal complaints and public inquiries branch. Her office put out a statement debunking the story in January 1990, but the flood continues unabated. Last month, the FCC was getting between 150 and 200 letters a day from computer network users incensed by the 'proposal'. The House of Representatives subcommittee on telecommunications still gets between 20 and 30 letters daily.

For the record, there was an FCC proposal in 1987 to impose something akin to a modem fee. But it was dropped after telecommunications services such as Compuserve mobilized a letter-writing campaign that dumped more than 10,000

complaints on the FCC. Over the past few years, the FCC has also been slowly implementing rules under the so-called Open Network Architecture regulations that could theoretically impose a fee for accessing computer network services. But the FCC has specifically created an exemption for such end-user network access, on the grounds that 'rate shock' would disrupt the rapidly growing computer telecommunications community.

No one seems to know where the latest round of rumours began, nor who started it. Andrew Blau, a telecommunications expert at the private Washington-based Electronic Frontiers Foundation, chalks it up to "cyberspace folklore". It is an electronic village version of the old 'urban myth', such as the baby alligator flushed into a New York City sewer that developed into a 20-foot man-eater.

Blau blames the persistent rumour on a misguided sense of civic responsibility on the part of some network user who finds an old warning when searching through databases of archived messages. The user may distribute the message (which typically consists of a short description of the 'proposal' and its dire consequences, a fill-in-the-blanks form letter and a list of congressional names and addresses) to other users, not realizing

that it is either five years old or simply a mistaken relic from the last round of rumours.

Although the 'facts' change subtly in each round of messages, most of them mention that the proposed modem fee was discussed on a San Francisco talk show and in the *New York Times*. No dates are ever given and, as far as the FCC has been able to determine, neither the article nor the talk show ever existed.

The FCC seems to be particularly vulnerable to rampaging rumour. Another legend that has made life miserable at the commission is that Madeline Murray O'Hare, the head of the US atheist movement, has petitioned the FCC to ban religious broadcasting. This story — which the FCC has explained countless times is utterly unfounded — has resulted in several hundred thousand letters.

To the computer network community's credit, each wave of speculation and warning is usually countered eventually by some user who actually bothers to make a few telephone calls and investigate before printing out the form letter and sending it off. One of those who have sought to stem the latest wave is David Dean, an anthropology graduate student at the City University of New York. As an unwitting propagator of the initial modem tax rumour, Dean felt an obligation to check out some of the sceptical replies he got from those who had seen it all before. After a call to Congress and the Electronic Frontier Foundation, Dean redeemed himself by uploading another message clarifying the situation.

According to FCC and congressional officials, the latest wave seems to be slowly dying out, no doubt due to efforts like Dean's. But it will, they predict, be back within a year. **Christopher Anderson**

EXPO site to become science park

Munich. When EXPO'92 shuts down in October, more than 200 hectares of pavilion-clad park land along Seville's river Guadalquivir will be turned into a science park. The scheme is part of a programme by the European Communities (EC) to support research and development in southern Europe.

The EC is donating 670,000 ECU (US\$850,000) to help develop the park in the Cartuja area of Seville. Much of the money will be used to facilitate the cooperative venture between three other EC-supported science and technology parks in the Mediterranean basin, at Montpellier (France), Bari (Italy) and Malaga (Spain). Experts from each of these sites will be pooling their experience to offer practical advice to the Cartuja Science Park, scheduled to open next spring.

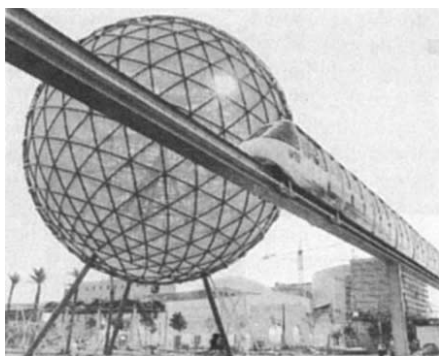
Juan-Carlos Moreno, whose company,

Sociedad Cartuje '93, coordinates future plans for the EXPO site, says that 20 companies, including Rank-Xerox and Cruzcampo, one of Spain's largest breweries, have expressed interest in the science park. They are now waiting for a government decision, expected within the next couple of months,

on financial incentives. At an annual land rental cost of 2,000 pesetas (US\$20) a square metre, the price is lower than comparable science and technology parks in northern Europe.

But cost is not the only incentive for prospective tenants. Dieter Eckert of Siemens, the German-based international electronics

company, says that the price of land has not played a serious part in its decision to locate in Cartuje. Already well-established in Spain, Siemens is eager to have a development centre there as well. **Alison Abbott**



The monorail passes the bioclimatic sphere at EXPO '92.

Max Planck stays put

Munich. The Max Planck Society has abandoned its 40-year-old plan to move to Berlin once the city regained its status as Germany's capital. At the society's annual meeting in Dresden, president Hans Zacher announced that the society, which runs major research institutes all over Germany, would stay in Munich "for the foreseeable future".

Plans to move to Germany's historical capital, first announced in 1952, reflected the ideological desire to occupy the home of the society's predecessor, the Kaiser-Wilhelm Society, in a reunited Germany. But as the cost of the recent reunification has delayed plans to move government offices to Berlin until 1997, the Max Planck Society has settled on legally registering itself in Berlin. **Alison Abbott**

NTT to unhook mobile phones

Tokyo. One of the world's biggest companies, Nippon Telegraph and Telephone Corporation (NTT), is preparing to spin off one of its most promising and most research-intensive businesses.

On 1 July, under a government policy aimed at promoting fairer competition in Japan's telecommunications industry, the rapidly growing mobile communications sector of NTT will become a new company, taking with it a sizeable and valuable section of NTT's army of researchers. But observers question whether NTT intends to divest itself fully of this part of its business and research enterprise.

NTT employs about a quarter of a million people and has an annual operating income of nearly ¥6.5 million million (US\$50,000 million). The company has more than 3,000 researchers in four centres and two laboratories doing fundamental work in such areas as optics, semiconductors, superconductivity and nanotechnology.

The success of the company is built on its monopoly of most of Japan's domestic telecommunications networks. In 1985, the Japanese telecommunications industry was privatized and several new companies entered the domestic market. But those companies have only a tiny market share because NTT still controls most of the telecommunications infrastructure.

In 1990, an advisory committee to the Ministry of Posts and Telecommunications recommended that NTT be split into regional companies, as occurred in the United States with AT&T, to promote fairer competition. But the recommendations were strongly resisted by NTT, some powerful leaders of Japanese industry and the Ministry of Finance, which has been making money by periodically selling off NTT shares to the public. In the end, it was agreed that only the mobile communications sector of NTT would be divested and that no further steps would be taken until at least 1996 (see *Nature* 344, 480; 1990).

Nevertheless, the loss of the mobile telecommunications sector would represent a very significant divestiture for NTT. Mobile communications companies are the "growth hounds" of the telecommunications industry, says David Hytha, managing director of Network Wireless Systems at AT&T Japan Ltd, as the use of automobile, portable and cordless telephones has soared in Japan and elsewhere in the developed world.

In the mid-1980s, Hytha says, few people in the telecommunications industry paid attention to wireless technology. Instead, all eyes were focused on high-speed digital cable networks and fibre optics. But now, he says, "everybody wants to go wireless". As a result, radio engineers and researchers on analog systems who were once considered

to be on the fringe have "suddenly moved to centre stage".

Japan trails the United States in research on mobile communications, Hytha says. It is especially deficient in radio engineers, mathematicians and software engineers capable of writing the rapidly evolving software and algorithms needed for mobile communications. The situation is made even more complicated by the existence of different standards for mobile communications in Japan, Europe and the United States.

The new company, to be called NTT Mobile Communications Network Inc., will take with it 250 mobile communication researchers from NTT. Or at least that is how it appears on paper. In fact, the researchers will not actually go anywhere. All of them will continue working at existing NTT research centres, 160 at NTT's Yokosuka research centre south of Tokyo and most of the rest at the Musashino research centre on the western outskirts of Tokyo.

NTT officials insist that the research and development arms of the two companies will be distinct. The researchers will be on different floors, says Shuichi Shindo, executive manager of the research and development department of NTT's mobile communications sector, and there will be no overlap in research between the two companies. An NTT spokesman says that the new company will be charged "at market rates" for access to NTT research and development.

But the division of research and development will not be clear-cut. Much basic research and development with applications in mobile communications, for example the development of integrated circuits, will be carried out at NTT's research laboratories. And some mobile communications researchers will stay at NTT, although they will not work on the development of portable telephone systems, cordless telephones or pagers and the technologies assigned to the new company.

Many observers question whether NTT intends to leave the field entirely. There are many emerging technologies that go beyond portable telephones, cordless telephones and pagers. A personal handy telephone, a cordless telephone that can be used outdoors, is expected on the market soon, and NTT may find it hard to turn its back on the potential of such technologies as wireless portable computers and wireless local area networks.

One thing is clear. NTT's departure from mobile communications will be very gradual. The new company will be owned totally by NTT, and, although NTT says it plans to follow government policy and reduce its holdings, shares in the new company will not be listed on the stock exchange until 1997.

David Swinbanks

IN BRIEF

London. Next month an international research institute will open at the University of Cambridge to strengthen collaboration between British mathematicians and their colleagues around the world. The Isaac Newton Institute for Mathematical Sciences is modelled on two centres in the United States — the Institute of Theoretical Physics in Santa Barbara and the Mathematical Sciences Research Institute in Berkeley, both part of the University of California.

The Newton institute will run four six-month research programmes a year in various cross-disciplinary subjects. The first batch will cover low-dimensional topology and quantum field theory, astrophysical dynamo theory, L-functions and arithmetic and the modelling of epidemics.

The impetus for the institute, according to its deputy director, Peter Goddard, came from a feeling that Britain was missing out on a good deal of international scientific activity. He and his colleagues could take part in cross-disciplinary projects in other countries, but were unable to host similar events.

Apparently they were right, for some 75 per cent of those who responded to a call for the initial programmes live outside Britain. Researchers will work at the institute for periods of between two weeks and the entire six months of a project.

I.M.

London. British staff at the Joint European Torus (JET), an experimental nuclear fusion project, staged the second of a series of one-day strikes yesterday (17 June) in support of a long-running campaign for parity of pay with their European colleagues. This follows the success of the first strike last Friday, which involved 112 British scientists and technicians and featured pickets outside the Culham Laboratory in Oxfordshire, where the project is based. A handful of European scientists joined the first strike. A ballot of union staff (see *Nature* 357, 270; 1992) showed 80 per cent support for action up to and including strikes.

Union representatives have had no response from JET's management to their complaints over pay or their concerns about what will happen to them after the JET project ends in 1996. Further strikes will be planned on a week-by-week basis.

I.M.

Munich. A three-day festival of science held last weekend in towns all over France attracted an estimated four million people. The 1,500 events ranged from science-fiction films to the Paris 'Market for the Year 2000' with its exhibition of future cars and cheese-making machines. The Ministry of Science and Space organized this first-ever national celebration of science and asked all French laboratories to open their doors to the public.

A.A.

Russian research may be put on lean diet

Moscow. Russia can no longer support its scientific community in the style to which it was once accustomed, according to Boris Saltykov, the science minister. And a new study suggests that the government is uncertain about whether to spend enough to maintain a first-class research enterprise.

Late last month, Saltykov presented to the Russian cabinet a proposal for a 'new science policy' that abandons support for across-the-board basic research and replaces it with an emphasis on priority projects. Moreover, the government will itself decide what those projects are. The leadership seems to have found the scheme attractive enough to appoint Saltykov deputy prime minister.

Under the new arrangements, there will be a stringent procedure for identifying which research institutions should retain their current privileges, including tax concessions and low rent. In future, funding for institutions will be replaced by funding for individuals, whose financial independence will be encouraged.

Saltykov's position is that Russia does not have the funds to satisfy everybody. Nor should it try, he says, for the future structure of the scientific community, the management of science and the mechanisms of funding will be very different from what

they are now. The new science policy will be implemented by two official bodies working in tandem: the Ministry of Science and the newly instituted Staff Office for Science and Technology Policies, the function of which is vaguely defined as "preparing and monitoring cabinet decisions".

This arrangement is by no means the loose and ineffectual organization it may seem. Saltykov and the head of the staff office, Ilya Lomakin-Rumyantsev, were successively the coordinators of a major study, under the aegis of the Russian Academy of Sciences, to analyse the status of research in Russia and to make recommendations.

Lack of money is not the only explanation for the stringency of the new policy. Lomakin-Rumyantsev says that the government is seeking deliberately to break away from the traditional organization of research under central authority. And while it is conscious of the hardships many scientists will encounter, it believes its new stringency will "encourage researchers to look for solutions on their own".

Lomakin-Rumyantsev says that there are three immediate tasks to tackle: to increase funding, not necessarily from the public sector, to change the mechanism by which that funding is disbursed and to "create a

solid and innovative legal framework for building a new model of science, largely based on nongovernment funds and deeply immersed in the production sphere".

The study on which Saltykov and Lomakin-Rumyantsev have been engaged has evidently taken a cold look at the international status of the new Russia. It says that Russia is now at a "bifurcation point" and that its choices will determine its scientific future. If Russia retains a position of leadership, the study argues, it will need an appropriately advanced scientific community. But if it emerges as a second-rate power, it will have to be content with a derivative research establishment to support production in high technology. The worst case, according to the study, is that Russia will join the 'Third World' and thus require no scientific establishment at all.

Although predictions are impossible, most analysts in Moscow believe that the compass needle is fluctuating between the second and third positions. But the adoption of new science policy suggests some optimism within the government about the future. "If we wanted to proceed along the third path", say Lomakin-Rumyantsev, "we would need no special policy at all."

Vladimir Pokrovsky

Debate over animals is given a personal touch

Washington. Sending thank-you cards to researchers across the United States from parents whose children are living because of medical interventions that were tested on animals is the latest wrinkle in a broad national campaign to win over the US public in a battle against those who oppose the use of animals in research.

Norma Myers Rumpf, with her healthy three-year-old son Lee asleep on her shoulder, spoke last week at a rally at the US Capitol about her volunteer effort to reassure scientists facing protests from animal-rights activists that their work is worthwhile. She received a standing ovation from representatives of a coalition of 250 organizations — universities, professional societies, patient support groups, health care companies, vol-

unteer health associations and individual researchers and physicians — formed six months ago to convey the message, as proclaimed on buttons and T-shirts, that "animal research saves lives". Rumpf presented Antonio Novello, the US Surgeon General, with 467 thank-you cards from the parents of children who, along with her newborn son, were saved from respiratory failure by the use of extra corporeal membrane oxygenation (ECMO), a procedure refined on sheep before it was applied to humans.

"Two out of three young people question the use of animals in research", said Frederick Goodwin, director of the National Institute of Mental Health, in reporting the results of a recent national poll on public support for biomedical research. "That tells you how much work we have to do in conducting an educational program on the value of research involving animals."

Although Goodwin says that he is confident "we will prevail in the long run", the rally reflected the concern among the biomedical community about the visibility and continued strength of the animal rights movement. "The 1980s was a lost decade for us", says Adrian Morrison, a veterinary researcher at the University of Pennsylvania who has been a target of protesters and a leader in the

effort to marshal public support. "It took scientists a long time to get organized, but I think that we've turned the corner."

For Rumpf, the turning point came two years ago when she read about a protest at the US National Institutes of Health (NIH) staged by People for the Ethical Treatment of Animals (PETA). "I thought about how hard it must be for a scientist to get up and go to work, knowing that people will be screaming in his face as he drives into his lab." The next day Rumpf wrote a letter of support to the NIH, and last year she presented more than 100 letters from thankful parents to Bernadine Healy, the director of NIH.

Rumpf, who works from her home in Arlington, Virginia, collects the letters and snapshots from grateful parents and sends them to scientists identified by various research organizations as working with animals. Her goal is to mail a thank-you card to every laboratory in the country, although she admits candidly that she does not know how long she can continue. Still, she seems willing to lick stamps for a while longer.

"There's nothing that I can do to repay them" for saving her child's life, she says. "The least I can do is pat them on the back and tell them to keep up the good work."

Jeffrey Mervis



Rumpf delivers cards to Surgeon General Novello.

Science popularity stakes

SIR — You ask “Is molecular biology yet a science?” (*Nature* 355, 201; 1992). I would answer “No, it’s rather a game of molecular roulette”. However, the real question is whether we do want molecular biology to become a quantitative science, and relegate it back to the ivory tower, or whether it should acquire ‘popular science’ status.

Anyone old enough to have witnessed its birth in the early 1960s can only agree with your characterization of molecular biology as “... well on the way to becoming a largely qualitative science”. Every older biochemist, steeped in the law of mass action and other thermodynamic principles applicable to biology, can add other examples for your list. All would show that neglect of quantification in molecular biology may cause important problems to be overlooked.

The new style molecular biology is also helping the literature to mushroom. A plethora of new journals with titles certain to contain elements of any of the words “molecular”, “genetic”, “biology”, “technology” or their permutations has appeared, and established journals are undergoing name changes to keep up. Witness finally the spectacle of learned societies such as the venerable American Society of Biological Chemists undergoing the ultimate humiliation of becoming the American Society of Biological Chemistry and Molecular Biology. How much more molecular is the biology done by these biochemists?

To be sure “... the triumphant pervasiveness ...” of molecular biology new style should also be credited for awakening the public to its promise. Never mind that those awakened are also the politicians, bureaucrats, science managers, conference organizers, and so on, always on the prowl for buzzwords. Thus, the adjective “molecular” has become indispensable to the young biologist seeking to enhance his or her career. The molecules and the biology (and their quantification) unfortunately often take second place.

But for the future of molecular biology perhaps the best solution would indeed be to let the de-quantification process take its due course, in order that the roulette aspects be allowed to take their predominant place. The defining moment for molecular biology to become a truly ‘popular’ science would then be its relabelling as ‘molecular biotechnology’, with “mobi-” or “mobiotech” as popular acronyms for short. That should ultimately result in a kind of ‘trickle-up’ process and perhaps improve the funding for the purists among us who wish to pursue their law of mass action ‘mobiotech-style’ to the hilt, and everybody would be happy.

J. M. Kaper

*Microbiology and Plant Pathology
Laboratory,*

*Plant Sciences Institute,
Beltsville Agricultural Research Center,
Beltsville, Maryland 20705, USA*

Counting on taxonomy

SIR — Gaston and May’s¹ estimate of the worldwide distribution of taxonomists among taxa and their bearing on the documentation and conservation of biodiversity can be made more precise for the taxonomists of Spain. Since 1987 we have conducted a continuous survey of Spanish taxonomists providing a census on an automated database system called DIRTAX^{2,3}. A total of 1,156 taxonomists has been recorded and their distribution by different categories is given in the table.

Strikingly, the percentage of taxonomists working on plants, microorganisms and animals closely fits estimates of US taxonomists. Distribution for lower categories of Kingdom Animalia are less concordant, especially the high number of arthropod taxonomists.

Approximately 60 per cent of registered taxonomists in Spain are professional, and 16 per cent have a temporary status, relying on grants.

The distribution of taxonomists by environment shows a preponderance of

DISTRIBUTION OF SPANISH TAXONOMISTS (%)

By kingdom	By environment	By status
Monera 3.6	Terrestrial 53.3	Professional 56
Plantae 28	Marine 16.6	On grants 16
Animalia 68	Freshwater 10.7	Amateur 13
Tetrapods 13.4	Fossils 10.2	Other 13
Fish 7.0	Parasites 2.8	
Arthropods 48.2	Subterranean 0.4	
Other inv. 31.4		

terrestrial taxonomists, despite the fact that the highest diversity is in the marine environment. Number of species in a phylum is highly correlated with number of taxonomists in that group, ($R^2=83.2$).

Elisa Bello

Jose M. Becerra

Antonio G.-Valdecasas

*Museo Nacional de Ciencias Naturales,
José Gutierrez Abascal,
2. 28006 Madrid, Spain*

1. Gaston, K. L. & May, R. M. *Nature* 356, 281–282 (1992).

2. G.-Valdecasas, A., Bello, E., Reyes, J. & Becerra, J. M. *Sci. Software* 5, 303–308 (1989).

3. Bello, E. & G.-Valdecasas, A. in *Abstr. 4th Int. Congr. Systematic Evol. Biol.* (1990).

A new kind of orphan drug?

SIR — In the United States an Orphan Drug Act was passed in 1983 to give companies incentives in the form of grants, tax credits and a seven-year marketing monopoly for the development of drugs for rare diseases — those that affect fewer than 200,000 people. Meanwhile some of these so-called orphan drugs, for instance aerosol pentamidine, which is used to treat pneumonia caused by *Pneumocystis carinii*, have made their developers hundreds of millions of dollars¹.

We should like to draw attention to a new species of orphan drug: time-proven, essential, standard drugs, unfortunately too cheap from the distributors’ point of view to justify further marketing.

In Austria, taeniasis (beef tapeworm) is endemic. Between 1986 and 1991, 82 cases have been verified in an Innsbruck laboratory serving a population of 900,000. At the time of the diagnosis of taeniasis, approximately 27 per cent of the patients had already undergone inappropriate anthelmintic therapy (mostly mebendazole) without clinical response².

Niclosamide is the drug of choice for *Taenia saginata*³. Since November 1983, niclosamide (Yomesan, Bayer) has no longer been registered in Austria. The company concerned says that the costs of official registration and distribution make continued marketing unjustifiable.

Praziquantel³, the only other drug effective in taeniasis, has been registered in Austria only for administration to animals. It appears that only animals promise sufficient consumption of the drug to warrant the costs of official registration and marketing.

This situation seems particularly ironic since niclosamide is listed in the WHO’s Model List of Essential Drugs⁴. Perhaps time-proven drugs that are too cheap and needed in too small quantities to ensure continued profit to the pharmaceutical companies should be designated as orphan drugs, to contribute to their more general availability.

F. Allerberger

M. P. Dierich

*Institute for Hygiene,
University of Innsbruck,
Innsbruck 6020,
Austria*

1. Gibbons, A. *Science* 248, 678–679 (1990).

2. Allerberger, F. *Mitt. Osterr. Ges. Tropenmed. Parasit.* 9, 65–72 (1987).

3. *Med. Let.* 34, 17–26 (1992).

4. WHO Expert Committee *WHO Tech. Rep. Ser.* 796 (1990).

5. Gershon, D. *Nature* 345, 759 (1990).

Science and sentimentality

D. S. Butterworth

After years of research, the scientific committee of the International Whaling Commission is about to put forward a sound method for setting catch limits for whaling. But will other considerations undermine the science?

THE convention governing the International Whaling Commission (IWC) states that its "regulations with respect to the conservation and utilization of whale resources . . . shall be based on scientific findings". The commission's main reason for introducing a moratorium on commercial whaling 10 years ago was the argued inadequacy of the scientific database for harvesting whale populations without exposing them to undue risks (*Nature* 357, 350; 4 June 1992). In response, the IWC's scientific committee has been trying to develop a revised management procedure (RMP) to provide adequate safeguards for stocks against the inevitable uncertainties associated with assessments of the sustainable yield which a marine resource can provide. The scientific committee looks set to present the IWC meeting in Glasgow this month with a finished product: an RMP for setting catch limits for whales, which has been more thoroughly researched than any comparable system for fisheries.

Many member countries of the IWC may find this outcome less than welcome. While the RMP was being developed, those countries eager to resume operations could easily be told to wait until scientific concerns had been properly addressed. Will a hidden agenda now surface — that opposition to resumed commercial whaling is not really a conservation but rather an 'animal rights' issue, backed by powerful public pressure groups opposed to *any* killing of certain 'special' animals, such as whales? This is a difficult viewpoint to argue because it involves ethical values, and whaling countries can object that others have no right to force their value judgement on them. Thus, until now, the scientific argument has provided a convenient front, because nobody would object to the standpoint that safeguards were needed to ensure that harvesting would not threaten whale populations with extinction.

The real debate in the IWC has been between some countries wishing to preserve industries, employment and a food source based on whales, and others wanting these animals classed as sacrosanct. The terms of the convention have required that this debate be conducted in a scientific guise, so that these hidden agendas have had to be played out in the scientific committee.

Whaling countries, angered by a moratorium decision which they considered to be without scientific justification, responded by commencing 'scientific whaling'. In terms of the rules of the convention, Iceland, Japan and Norway were entitled to issue permits to their nationals to catch some whales for the purpose of scientific research. As far as the lethal components of these research programmes are concerned, the empirical evidence thus far of their value towards an improved basis for management is poor. However, the sighting survey component of this research to provide estimates of abundance is agreed to be essential. Potentially the strongest defence for 'scientific whaling' is that because surveys are enormously expensive, it is not unreasonable to recover the costs through harvests that are sufficiently low to pose no risk to the stock. But the countries concerned have never argued this, perhaps because they preferred the combination of a legally stronger, but scientifically weaker, case.

The line of attack by countries opposed to whaling was to emphasize scientific uncertainty. But as the scientific database for whale populations improved during the 1980s, particularly as sighting surveys started to provide reliable estimates of population sizes, the uncertainty goalposts were shifted to try to continue to get stocks declared 'protected'. In terms of an earlier management approach adopted by the commission in 1975, a stock would be accorded 'protected status', with no catches permitted, if its size fell below 54% of the pre-exploitation level (K). However,

once data became available which, in terms of previously accepted approaches, would have indicated that the stock was definitely above this protection level of 0.54 K , some members of the scientific committee would obfuscate or insist that stricter criteria had to be met before this conclusion could be drawn. The United States, for example, was in a difficult position during the 1980s over the catches of bowheads desired by Alaskan eskimos under the IWC's aboriginal whaling scheme, as this population was considered to be severely depleted. In 1987, general acceptance in the scientific committee of a lower bound of 1% for the MSY rate* removed scientific obstacles to the eskimos' request, thus freeing the United States from a possible encumbrance when mounting an attack in the IWC on the 'scientific whaling' programmes of the whaling countries. Yet subsequently, some members of the committee would accept only 0% as a lower bound for minke whale assessments. When even this was insufficient to have the minke whales off Iceland declared 'protected', further factors were introduced into the debate to shift the goalposts still further. Eventually, Iceland lost patience and decided to leave the commission.

The term 'protected status' has been skilfully used to give the public the impression that a stock placed in this category would be in danger of extinction if any catches were taken from it. In fact, the original selection of the 0.54 K value was related to catch-maximization considerations, and had nothing to do with any likelihood of extinction.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Japanese factory ship *Nisshin Maru* hauls a whale up the slipway.

The RMP has been constructed in a way that attempts to take specific account of these 'uncertainty' concerns. Normally, catch limits in fisheries are calculated using the 'best' mathematical model of the resource. In contrast, the RMP concentrates on the risks to the population that could emerge if this 'best' model is incorrect. A different model, also consistent with the available data, may better represent the real situation. The RMP thus seeks to set catch limits in a 'robust' manner: whatever the (unknown) real situation, the catches to be taken should neither waste the resource nor put it at risk.

The Achilles heel of this approach to problems such as the design of an RMP is that it is impossible to develop a procedure that is robust to every imagined possibility. A restricted list of plausible alternatives has to be agreed, but such a process can easily be frustrated if bounds are not placed on speculation, which can sabotage attempts to achieve the consensus necessary in the IWC's scientific committee.

It might seem that firmer leadership of the committee is all that is needed to prevent speculation that is not constrained by data, and to stop members switching viewpoints without proper reasons. But the chairmen of the scientific committees of international fishery organizations can be severely limited by the political climate in which they have to operate and attempt (frequently without success) to achieve consensus. The scientists who serve on these committees are delegates of the countries that are members of the organization concerned, and are often employees of their governments. (Some governments even appoint people without scientific credentials to such committees, at times instead of those who have!) On occasions these scientists are under instructions to argue in support of a position adopted by their government for reasons other than science alone (for example, economics, or 'animal rights' concerns), and also to ensure that their arguments are reflected in the scientific record. Indeed, how realistic is it to expect a country to appoint a scientist to such a committee, if he may present analyses with conclusions inconsistent with his own government's stance? Fortunately, not all countries operate in this manner, and the United States in particular demonstrates a pleasing (if not always perfect) recognition of the rights of its scientists to express their views independently of

political concerns. The IWC scientific committee also makes much greater use than is the norm of 'invited participants' — specialists who are not members of a national delegation — but intends their role to be technical only.

The role of scientists is not to make the final management decisions — that is a job for the politicians. Scientists' responsibilities are to provide the politicians with predictions of the likely results of alternative management options, so that their decisions are made on an informed basis. These predictions need to be made on as objective a scientific basis as possible: but how likely are the scientific committees of international

are assessed to be below 0.1K. Even making allowance for possible inadequacies in the spawning-biomass-per-recruit basis for these assessments, and also for indications that fish stocks are probably most productive at smaller fractions of K than are whales, a gross inconsistency still remains. This is not to suggest that the United Kingdom and the United States are unconcerned about the status of these resources, or that they are not attempting some remedial action: but nevertheless, they (and others) are clearly not prepared to take the action they demand from countries with interests in whaling, when their own fisheries (and socio-economic interests) are concerned.

Part of the responsibility for this problem lies with the fisheries science community, whose thinking on harvesting strategies remains too rooted in the deterministic concepts of earlier decades. A key factor in selecting between alternative strategies is the risks to the resource that each involves. But risk has to take account of stochastic aspects, and fisheries scientists have been dilatory in adopting definitions and norms that will provide a basis for the consistent approaches to fisheries management that the earlier deterministic benchmarks (such as the $F_{0.1}$ strategy) were intended to achieve. Thus, one should perhaps be more forgiving if many member nations of the IWC, having agreed a risk-related objective for the RMP which requires no serious increase in the risk of extinction of a stock through exploitation, labour under the misapprehension that a 5%

chance of a small catch from a stock slightly below 0.54K constitutes a serious risk that the stock will be rendered extinct. The general application of equally risk-averse criteria to all marine fisheries would necessitate immediate closure of the overwhelming majority.

The time seems overdue for scientists to speak out against the near-farcical pronouncements of some international organizations regarding endangered species. The Antarctic minke whale is agreed to number a precisely determined $760,000 \pm 9\%$, which compares with a cumulative historical catch from this resource of only 115,000. Viewed as a whole, its status is near the best, and best-determined, of any of the world's marine resources. Nevertheless, the Convention for International Trade in Endangered Species (CITES) lists minke whales as a species in danger of extinction. By contrast, a proposal at last March's CITES meeting to list the western North Atlantic bluefin tuna similarly was successfully opposed by a number of nations (including the United States),

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Polar eskimos flensing a narwhal.

fishery organizations to achieve this, given the manner in which they are constituted? Is there any real role for science, or does the IWC example serve only to show how vulnerable science is to reduction to irrelevancy, given hidden agendas such as economic considerations or animal rights?

I believe that scientists, particularly those independent of government agencies, still have a very important role to play in striving for consistency. The reason that adoption of the RMP was delayed at the IWC meeting in 1991 was that some countries, including the United Kingdom and the United States, expressed concern at probabilities in excess of 5% that catches might be taken from a stock which, though assessed to be above the 0.54K protection level, was actually below this. Yet neither the United Kingdom nor the United States has proposed sanctuaries or imposed moratoria on the fishing of cod in the North Sea, or cod and flounder off the north-east US coast, despite the fact that the spawning biomasses for these resources

* The MSY rate is a measure of the potential productivity of a resource. It is the ratio of the annual maximum sustainable yield to the population size at which that yield is obtained (conventionally taken to be 0.6K for baleen whales). Evaluations of risks associated with the RMP have all been based on an MSY rate of 0.66%. However, the appropriateness of this choice is coming under increasing question, as all recent evidence points towards baleen whale stocks having MSY rates of about 3% or more.

despite generally accepted assessments that the spawning biomass of this resource has been reduced to less than 0.2K and is probably below even 0.1K. I do not intend to suggest that a moratorium is necessary for this tuna fishery; other measures than a total suspension of fishing can serve the needs of conservation adequately. My purpose is to point again to the gross inconsistencies in attitudes to risk which many nations show in the management of fisheries.

The real problem is that these issues involve questions of 'value', rather than of science and the consistency which it demands. 'Animal rights' lobbyists, or at least their public supporters (and financiers), attach much greater 'value' to a whale or a seal than to a tuna or a cod. That may be a perfectly defensible position, but what becomes dangerous is the use of alleged scientific concerns as a surrogate rationale for this standpoint. This serves only to devalue science in the eyes of the politicians, who have to take particular account of public pressure in making decisions. If science is manipulated or rendered irrelevant on issues where public emotion can be tapped, it will simultaneously become redundant on less emotive environmental issues, to the detriment of long-term conservation.

Scientists therefore have a responsibility to emphasize the differentiation of scientific issues and value judgements. New Zealand, at least, is honest enough to state that its opposition to resumed commercial whaling is based on such value judgements. In contrast, I have less sympathy for the position of 155 scientists reported to have signed a "statement of scientific findings regarding commercial whaling and science", circulated by the Cetacean Society International. They state therein that they find that "the commercial killing of whales and other cetaceans unnecessarily interferes with scientific research. Its disruption of the marine ecosystem needlessly conflicts with the objective study of the natural world", and go on to "therefore recommend an end to all commercial whale killing, for the foreseeable future". If there was any scientific validity in this argument, would it not require moratoria on aboriginal whaling, and indeed also for all marine fisheries? Or is this just another example of science being confused with value judgements outside the scientific domain; and a more depressing one at that, as scientists themselves are involved?

All being well, the IWC's scientific committee will present the commission meeting in Glasgow with an RMP, together with the catch limits which it indicates for the Antarctic and North Atlantic minke whale resources. These catch limits will be allocated to much smaller areas of ocean than was pre-

viously the case — this enforced wide distribution of any catch has been the solution adopted in the RMP to the perennial problem of uncertainty about the appropriate placement of stock boundaries for marine resources. The committee will also advise that the RMP is likely to allow catches when the true population abundance is below the protection level of 0.54K only if the MSY rate is low (near 1%); but that even then, the size of these catches would be too small to make any marked impact on the recovery rate of the stock.

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Bluefin tuna — endangered species?

What will the IWC do with this information? Doubtless the question of the humaneness of the methods used to kill whales in commercial operations will be raised. But countries in which there are aboriginal whaling activities (for which the death times are much longer than in commercial operations), may attempt to play down this issue.

One possible outcome is that the commission accepts the scientific committee's RMP and associated specification of catch limits for resumed commercial whaling of minkes in the North Atlantic and Southern Ocean. The RMP adopted will be unnecessarily conservative, with the commission's confusion over the pertinence of a 0.54K 'protection level' to the risk of extinction having resulted in the choice of a procedure that will be unable to realize the potential yield from stocks with MSY rates much above 1% (contradicting another of the commission's agreed objectives for the RMP). Whaling countries may be prepared to compromise on this aspect in return for a majority vote to lift the moratorium.

Anything less than this will probably lead, later if not immediately, to Norway and Japan joining Iceland in withdrawing from the IWC. These countries may

in due course attempt to expand the recently formed North Atlantic Marine Mammal Commission into a 'new' IWC, arguing that the existing body has become a "commission for the prohibition of whaling". The animal-rights lobby might not be unhappy with such an outcome, surmising that their goals might be more easily attained by encouraging consumer boycotts of goods from the 'rebel' countries, than by relying on nominally scientifically based decisions at the IWC.

The recent and sudden French proposal for a whale sanctuary in waters south of 40° S to aid the "rehabilitation of the Antarctic marine ecosystem" is a wild card. Under the RMP, the severely depleted whale species of the Southern Ocean, such as the blue and fin whales, will remain completely protected for many decades. Scientifically speaking, the proposal — a weak and thinly disguised attempt to inhibit harvesting of Antarctic minke whales under the RMP, despite its protestations to the contrary — suggests desperation among the 'animal rightists'. But it could be an astute initiative politically, as a major concern of many member countries of the IWC that have no current interest in resumed whaling themselves is short-term public relations, to which scientific consistency takes a poor second place. In these terms alone, such countries may see voting against any sanctuary proposals only as a net loss to themselves.

This illustrates the real problem that the IWC faces in trying to achieve a balanced approach to the possible lifting of the moratorium. There is no pressure to reach a compromise when some parties to the decision see no gain for themselves in allowing others to recommence commercial whaling. Much will depend on the line taken by the United States, where there is legislation that would 'require' the US government to take punitive action against countries that might leave and commence whaling activities outside the IWC. Although 'animal rightists' might see such a situation to be in their longer-term interests, the US government may not share their enthusiasm at being placed in such a position, and may therefore be anxious to secure some compromise.

Whatever happens, fisheries science will not be left without some benefits. Controversy is a useful stimulant to scientific endeavour, and the process followed by the IWC's scientific committee to develop an RMP to take direct account of concerns about ever-present uncertainty can be duplicated with advantage for the management of many other of the world's fisheries. □

D. S. Butterworth is in the Department of Applied Mathematics, University of Cape Town, Rondebosch 7700, South Africa.

Save the beaches, not the buildings

Canute, the eleventh-century king of England, is best known for his demonstration that even he could not hold back the tides. Can the US Congress learn this lesson?

ALONG most of the coastline of the United States, the beaches are eroding. It is nature's way. The coastal barriers, narrow strips of sandy land and wetlands that mark the East and Gulf coasts, stretching more than 15,000 miles from Maine to Texas, are particularly vulnerable to erosion. As geographer Rutherford H. Platt of the University of Massachusetts has written, "Coastal barriers are naturally unstable landforms whose exact position, width, length and topography are subject to gradual and sometimes very rapid change. A striking characteristic of most coastal barriers in their natural state is their tendency to migrate or recede gradually landward."*

That being so, it hardly seems sensible that people build houses on shifting sands. Perhaps that can be explained by man's romantic love of the sea. But what can possibly explain the fact that the US Congress years ago agreed to provide flood insurance, at federal expense, for home owners whose beachfront property is damaged by storms and erosion? Congress's outrageous propensity to bend to the wishes of special interests is the only plausible answer because nothing in the scientific literature suggests that 20-storey luxury condominiums and the Cape Cod seashore were meant to be part of the natural order of things.

Legislation that passed the US House of Representatives a year ago and is now under consideration in the US Senate would all but eliminate federal subsidies for flood insurance and also limit the federal government's commitment to restore shifting sands. Real estate investors, who stand to gain the most when apartments or colonies of houses are built on the beach, are up in arms, arguing that property values would drop significantly if property owners had to pay for their own insurance. It is a prospect that brings tears to the eyes.

The wonder is that legislation to relieve the government of this foolish obligation has been so long in coming. It is not as if coastal erosion were a new phenomenon, or one that man had learned to control. Quite the contrary.

Nearly a century ago, a hurricane pounded the Gulf coast at Galveston, wiping out scores of buildings. The city's response? It built a seawall ten miles long. The wall is still standing but the beach eroded anyway. Nevertheless, people determined to dwell at the water's edge have continued to believe that seawalls can protect them. Platt recounts the effect of a series of seawalls along

the New Jersey beaches. "Armoring often leads to the loss of the beach itself," he says, "due to increased wave scour, steepening of the shore profile and loss of dunes as a source of beach sand."

Another way to thwart nature, often equally ineffective, is through the technique of 'beach nourishment'. Sand from a nearby inlet or sandbar is pumped onto the eroding beach. A nourished beach on Long Island washed away within a year, at no small cost.

Four years ago, Ocean City, Maryland, dumped 2.4 million cubic yards of sand on its disappearing beaches, which simply eroded again under battering by severe storms. But Ocean City did not give up. It just dumped more sand — 3.6 million cubic yards the second time around at a cost of \$45 million, but Congress's General Accounting Office predicts that costs will easily reach \$250 million by the end of the decade because the seven-mile beach will almost certainly have to be renourished with millions of cubic yards of sand every couple of years. Dredgers pull sand from the ocean and send it through steel pipes to the beach so that vacationers can lie in the sun (and increase their chances of getting melanoma).

In a report issued in 1987 by the US National Academy of Sciences (*Responding to Changes in Sea Level*), academy scientists estimated that a predicted rise of sea levels, to a potential high of 3.5 metres 20 years from now, will increase the likelihood of erosion, as will global warming that will lead to more hurricanes and, in turn, more coastal erosion.

The proposed legislation also takes into account problems of erosion along the shorelines of the Great Lakes, which are less vulnerable but nevertheless threatened. According to Platt, "The Lakes reached a record low level in 1964 and then rose to two successive peak periods in 1973–74 and 1985–86, with a difference of nearly six feet between the lowest and highest levels." During high water, hundreds of miles of coastline receded to the detriment of homeowners who built too close to the water's edge and taxpayers who had to bail them out.

Environmental groups, including the Sierra Club and the Audubon Society, have lobbied hard for legislation that would point the way to sensible coastal conservation. Even the Federal Emergency Management Agency (FEMA), a federal agency best known for its plans to evacuate cities in the event of nuclear war by directing thousands of motorists on to already clogged highways, has put what little weight it has behind

the pending legislation.

The legislation would not only gradually get the US government out of the business of paying for flood insurance for beach dwellers, it would also put strict limits on where new buildings could be constructed, based on academy estimates of the pace of erosion at various places along the coast.

If it is true that the extent of erosion can be predicted over 10 years, 30 years and even 60 years, as academy scientists believe, laws can be rewritten to take anticipated erosion into account. For instance, no one would be permitted to build anything new in a 10-year zone, while those who already have beachfront homes would be given an insurance allowance for either demolishing or relocating their houses. And substantial limits would be placed on new construction in 30-year and 60-year zones.

Needless to say, the free market effects of this legislation would bring property values down, just as landowners fear. But that is what the free market is all about.

This all makes very good sense. Man-made destruction of the coastline would be thwarted and the taxpayers from the cornbelt would stop supporting eastern beach lovers. The evidence is that destructive building habits, especially the miles of high-rise buildings that mar the East coast from New Jersey to Florida (it is no longer possible even to see the ocean from Florida's ocean highway), were fostered in the first place by the 1968 laws that made federally subsidized insurance possible.

Today, the owner of beachfront property valued at about \$200,000 pays a mere \$950 a year for flood insurance. On the open market, a similar policy would cost more than \$18,000, assuming that private insurers would be willing to take the risk at all. New development would essentially come to a dead halt, reacting to the free market principles that the administration of President George Bush so favours. Nature would be the real beneficiary.

But it may be too much to hope for. The political influence of the real estate lobby, not to mention that of home owners who argue that a change in insurance laws would amount to an unconstitutional taking of land, is not to be underestimated, but the US Congress should approve legislation to limit federal insurance subsidies to people who want to build on eroding sands.

Barbara J. Culliton

* *Cosmos*, 1991. The annual journal of the Cosmos Club, Washington, DC.

Devious devices of *Salmonella*

Daniel A. Portnoy and Gregory A. Smith

THOSE working on host-parasite relations and on signal transduction will find common interest in the paper by Galán and colleagues on page 588 of this issue¹. The authors show that the enteric pathogen *Salmonella typhimurium* somehow stimulates the host cells' receptors for epidermal growth factor in order to invade those cells.

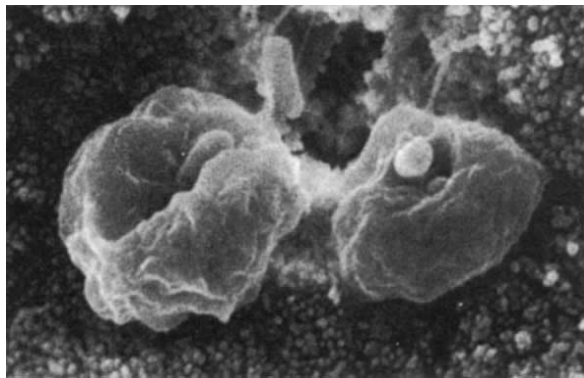
Many microbial agents of human disease attach to and invade host cells which are not phagocytic. The process of invasion — more accurately termed parasite-specified endocytosis² — involves the interaction of microbial determinants with host receptors capable of transmitting signals to the cytoskeleton which trigger entry of the agent. Galán *et al.* show that the entry of the bacterium *S. typhimurium* into a cultured epithelial cell is coincident with tyrosine phosphorylation of the receptor for epidermal growth factor (EGF). In addition, a bacterial mutant which still binds to host cells, but does not enter them, fails to induce the phosphorylation of the EGF receptor. Surprisingly, the mutants' problems can be solved by the addition of EGF — that is, with EGF, the mutant *S. typhimurium* is internalized. Lastly, both *S. typhimurium* and EGF induce a pattern of host proteins phosphorylated at tyrosine which is remarkably similar. Taken together, these observations suggest that the bacteria are directly or indirectly stimulating the EGF receptor.

The new work has its roots in the classic electron microscopic study by Takeuchi³, who documented events occurring at the intestinal brush border of infected guinea pigs. Most notable was the observation of local degeneration of the intestinal microvilli and the occurrence of blebs associated with bacterial internalization (see figure). Experiments *in vitro*, using cultured cells, have since found a direct correlation between the entry of *S. typhimurium* into cells and the profound reorganization of the host cell's cytoskeleton at the site of entry^{4,5}. But the molecular details of this example of parasite-directed endocytosis have remained obscure.

Entry of *S. typhimurium* into host cells seems to be rather different to entry of another model enteric pathogen, *Yersinia pseudotuberculosis*. In the case of *Y. pseudotuberculosis*, both binding and entry are mediated by a single domain of a bacterial surface protein called invasin⁶. Invasin works through interaction with

integrin receptors, and although these receptors do not normally mediate internalization, Isberg and colleagues have shown that they can do so when bound with high enough affinity. In the case of *S. typhimurium*, the nature of the bacterial ligand(s) is not known. But because *S. typhimurium* has genetically separate loci for adherence and invasion (unlike *Y. pseudotuberculosis*) a two-step model is likely — adherence to an as yet unknown receptor, and cellular activation through the EGF receptor.

The implied involvement of the EGF receptor certainly makes it the most



Scanning electron micrograph of polarized cultured epithelial cells, showing the formation of cytoplasmic membrane blebs which result from *Salmonella typhimurium* infection. (Courtesy of J. E. Galán.)

likely candidate as a direct target of host-cell activation caused by *S. typhimurium*. One attractive possibility is that the bacterium possesses an EGF analogue or EGF-like domain on a surface protein. The phosphorylation of the EGF receptor during invasion could however be mediated by an indirect process, such as crosstalk with another kinase. The stimulation of such an activity could trigger the second step in the *S. typhimurium* entry pathway as an upstream effector of the EGF receptor. Indeed, the EGF receptor does appear to be a target for transmodulation *in vivo*, and a substrate of other tyrosine kinases *in vitro*^{7,8}.

A major question arising from this study is the role of the EGF receptor in this process. Perhaps the receptor itself is internalized subsequent to activation, thus providing a direct route of entry for *S. typhimurium*. But it is known that EGF has dramatic effects on the host cell, including membrane ruffling, stimulation of pinocytosis, redistribution of cell-surface receptors (including internalization of its own receptor), generation of calcium flux and changes in cellular pH (ref. 8). Perhaps the induction of

macropinocytosis, which is known to occur in macrophages and other cells in response to growth factors^{9,10}, results in the nonspecific entry of bound bacteria.

However the authors feel that the effect is specific for *S. typhimurium*, as an adherent, but non-invasive, strain of *E. coli* was not internalized. Thus binding of wildtype *S. typhimurium* may induce a global change, mediated by the EGF receptor, which facilitates bacterial uptake by specific receptors not usually involved in phagocytosis. Alternatively, activation of the EGF receptor may result in an increase in the number of available receptors used by *S. typhimurium* by redistribution of these receptors from an intracellular pool. Such an increase could lead to bacterial entry, as shown by Tran Van Nhieu and Isberg.

Their evidence suggests that particles bound to integrin receptors, which are usually maintained at the cell surface, are internalized by an increase in receptor number (personal communication). These examples of receptor modulation are reminiscent of a story in which human monocytes bind complement-coated particles via the integrin receptor CR3 ($\alpha_m\beta_2$) but do not internalize the coated particles¹¹. However, phorbol-ester-stimulated monocytes or monocytes plated on fibronectin do mediate entry of the coated particles.

This is not the first example illustrating a connection between phosphorylation and microbial pathogenesis. Pathogenic yersiniae and vaccinia virus express tyrosine phosphatases which act on host proteins¹². The precise role of the phosphatases is not known, but in *Yersinia* it may be to prevent phagocytosis of bacteria bound to bacteriocidal macrophages. Thus by mediating the dephosphorylation of host proteins, the bacteria may avoid their demise. This is in contrast to *S. typhimurium*, which seems to stimulate phosphorylation to promote its own uptake.

It is becoming clear that tyrosine phos-

1. Galán, J. E., Pace, J. & Hayman, M. J. *Nature* **357**, 588–589 (1992).
2. Moulder, J. W. *Microbiol. Rev.* **49**, 298–337 (1985).
3. Takeuchi, A. *Am. J. Pathol.* **50**, 109–136 (1967).
4. Finlay, B. B., Ruschowski, S. & Bedhar, S. *J. Cell Sci.* **99**, 283–296 (1991).
5. Francis, C. L., Starnbach, M. N. & Falkow, S. *Molec. Microbiol.* (in the press).
6. Isberg, R. R. *Science* **252**, 934–938 (1991).
7. Koland, J. G., Maddison O'Brien, K. & Cerione, R. A. *Biochem. biophys. Res. Commun.* **166**, 90–100 (1990).
8. Rozengurt, E. *Science* **234**, 161–166 (1986).
9. Haigler, H. T., McKanna, J. A. & Cohen, S. *J. Cell Biol.* **83**, 82–90 (1979).
10. Racosin, E. L. & Swanson, J. J. *exp. Med.* **170**, 1635–1648 (1989).
11. Wright, S. D., Licht, M. R., Craigmyle, S. & Silverstein, S. C. *J. Cell. Biol.* **99**, 336–339 (1984).
12. Clemens, J. C., Guan, K., Biska, J. B., Falkow, S. & Dixon, J. E. *Molec. Microbiol.* **5**, 2617–2620 (1991).

phorylation is centrally involved in host signal transduction and communication with the cytoskeleton. Thus it should come as no surprise that microbial pathogens have evolved mechanisms to affect this pathway. In the near future this avenue of research should improve our understanding of both microbial

pathogenesis and eukaryotic pathways of signal transduction. □

Daniel A. Portnoy and Gregory A. Smith are in the Department of Microbiology, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104-6076, USA.

QUANTUM OPTICS

New spectroscopic frontiers

Stephen M. Barnett

SPECTROSCOPISTS may be able to squeeze improved sensitivity out of their spectrometers by using a novel state of light. E. S. Polzik, J. Carri and H. J. Kimble (*Phys. Rev. Lett.* **68**, 3020–3023; 1992) show how it is possible to reduce noise levels below the quantum limit using a new, tunable source of so-called squeezed light, surpassing even the best laser systems.

The introduction of lasers and the ensuing development of nonlinear optical techniques have revolutionized spectroscopy. This revolution has been due to the unique properties of laser light; it is coherent, intense and, of particular relevance to spectroscopy, it can be made highly monochromatic, stable and tunable. Moreover, laser light is very 'clean', in that it does not have the large intensity and phase fluctuations associated with conventional light. Yet even lasers do not represent the ultimate in clean light sources, as Polzik *et al.* show.

The question facing spectroscopists is how accurately they can detect the absorption of light. The answer depends upon both the degree of precision with which they can determine the level of light impinging on the sample and the efficiency with which they can detect the unabsorbed light. Polzik *et al.* tackle the problem in two steps. Frequency-modulated spectroscopy, a subtle trick of the trade to make the most of laser light, addresses the first step by concentrating on the differences in the absorption and phase shifts experienced by light at different frequencies, which are easier to handle than absolute levels at a single frequency.

The modulation of the laser light produces sidebands at frequencies either side of the original laser frequency. The differential absorption experienced by these sidebands converts the frequency modulation to amplitude modulation. This is detectable in the beat note between the sidebands and the carrier at the original laser frequency.

But the sensitivity of this technique is limited by the precision with which one can ensure that the two sidebands have the same intensity. Laser light can be

made highly coherent and stable, but even after all sources of technical noise have been removed there remains an intrinsic quantum noise. This residual noise is responsible for poissonian fluctuations in the detector photocurrent. Using this photocurrent to infer the intensity, one finds an uncertainty that is proportional to the square root of the mean intensity. The precision with which one can specify intensities or intensity differences is ultimately limited by this uncertainty — the standard quantum limit. Frequency-modulated spectroscopy with lasers can attain but not exceed this limiting sensitivity, which is where squeezed light comes in.

The origin of the standard quantum limit lies in the quantum nature of the electromagnetic field. In quantum optics the sinusoidal and cosinusoidal oscillating components of the field are incompatible. This is analogous to the incompatibility of the position and momentum of a particle in quantum mechanics. Heisenberg's uncertainty principle places bounds on the precision with which we can discern both the position and momentum of the particle. It is possible to improve one's knowledge of the particle's position but only at the expense of losing information about its momentum. In the same way spectroscopists are restricted in the extent to which they can determine both components or quadratures of the electromagnetic field. All field states are similarly restricted, even the vacuum. The uncertainty in the field quadratures is the origin of the quantum fluctuations responsible for the standard quantum limit.

Laser light has an equal level of uncertainty in both quadratures of the field. However, squeezed light has one quadrature more precisely specified at the expense of increased uncertainty in the other. The level of uncertainty in the reduced-noise field component or squeezed quadrature is even less than that associated with the vacuum. When squeezed light is superposed on laser light, the resulting beam has reduced intensity or phase fluctuations, depending upon the phase of the original laser

beam. If the laser light is in phase with the squeezed quadrature then the resulting beam has reduced intensity fluctuations. But if they are out of phase, the level of intensity noise is increased and the phase fluctuations are reduced.

Polzik, Carri and Kimble exploit this suppression of quantum noise as the second element in their experiment. By generating squeezed light with reduced intensity fluctuations they have shown that it is possible to increase the signal-to-noise ratio beyond the standard quantum limit. The lower level of intensity uncertainty means that small changes of intensity due to weak absorption can be measured more accurately. Moreover, by reducing the background quantum noise level, weaker spectroscopic features may become discernible.

Central to spectroscopic applications is the requirement for tunable light. Polzik *et al.* have developed the first tunable source of squeezed light for their experiment. Its heart is a powerful titanium-sapphire laser (itself a recent development) which produces blue light. This light is passed through a crystal of potassium niobate in a resonant optical cavity, where some of its photons are converted into photon 'twins' of half the frequency. These make the tunable squeezed light at a wavelength of around 855 nm, which has photocurrent noise fluctuations 5 decibels (about a factor of three) below the standard quantum limit. This level of squeezing is retained as the source is tuned over 2 gigahertz. The squeezed light is mixed with frequency-modulated laser light to give a probe beam with reduced quantum noise. Squeezing the quantum noise associated with the sidebands reduces the level of noise seen in the beat note (and so improves the signal-to-noise ratio) by more than a factor of two.

Enhanced spectroscopic sensitivity is not the only benefit offered by squeezing light. Reducing the level of phase fluctuations instead of the intensity noise can improve the sensitivity of interferometric measurements. The ability to reduce the noise level of modulated light, central to the new development in frequency-modulated spectroscopy, has potential in optical communications. Squeezed light can modify fundamental radiative processes such as spontaneous emission. It might be possible to induce ultranarrow spectroscopic features by detecting the decay of atoms with squeezing-reduced natural linewidths. Tunable squeezed light is the key to unlocking a new frontier in spectroscopic precision. □

Stephen M. Barnett is in the Department of Physics and Applied Physics, John Anderson Building, University of Strathclyde, Glasgow G4 0NG, UK.

The box and the rod

Peter Parham

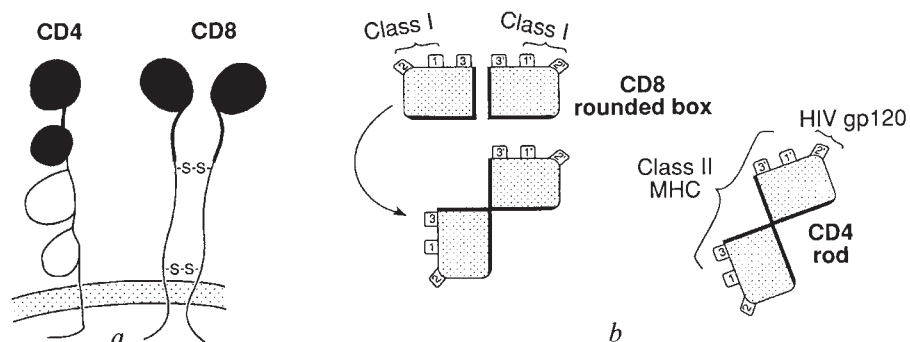
T-CELL receptors recognize peptides bound either to class I or class II major histocompatibility complex (MHC) molecules. Abetting these relationships are the coreceptors, CD4 and CD8, which respectively engage conserved regions of class I and class II molecules and are mutually exclusive in their expression by mature T cells. Structural similarities between class I and II MHC molecules, and the symmetry of their interactions with T-cell receptors and coreceptors, initially suggested that CD4 and CD8 would be very closely related molecules and show but minor variations upon a theme. Such notions were challenged by the considerable differences found in the primary structures of CD4 and CD8 polypeptides. Now, the crystal structure of CD8 (ref. 1) and that previously described for CD4 (refs 2 and 3) permit comparison of the MHC-binding sites of the two coreceptors. Although the differences between them are significant, so too are the similarities.

CD4 and CD8 are membrane glycoproteins with much of their mass hanging out of the cell. CD4 is a single polypeptide with an extracellular part consisting of four immunoglobulin-like (Ig-like) domains. CD8 is a dimer of polypeptides — either α -chain homodimers or $\alpha\beta$ heterodimers — each having an amino-terminal Ig-like domain followed by an extended and heavily glycosylated region. For both coreceptors crystallography was performed on fragments of the protein implicated in the interactions with MHC molecules and, in the case of CD4, with human immunodeficiency virus (HIV). The fragment from CD4 consisted of the two amino-terminal Ig-like domains^{2,3}, that from CD8 of a dimer of the amino-terminal Ig-like domain of the α -chain (*a* in the figure).

From sequence comparisons, the amino-terminal domain of the CD8 α -chain was predicted to resemble most closely an Ig-variable region⁴, and Leahy *et al.*¹ find the three-dimensional structure is "identical to that found in immunoglobulin variable domains with two antiparallel β -sheets of five and four strands connected by a conserved topology of loops". Furthermore the dimerization and relative orientation of the two domains is similar to that between the variable regions of antibody heavy and light chains or the two light-chain variable regions in a Bence Jones protein. To Leahy *et al.*, "The overall shape of CD8 is more like a box with rounded edges than, for example, a sphere"¹.

Does this mean that CD8 binds class I

MHC molecules in a manner analogous to an antibody, that is with the loops corresponding to the complementarity determining regions (CDR)? Yes, according to Sanders *et al.*⁵, who showed that mutations in the CDR1- and CDR2-like loops of CD8 α interfered with the binding of class I MHC molecules. The site of CD8 interactions has been mapped to α_3 , the domain of the class I heavy chain resembling an immunoglobulin constant domain, and specifically to an exposed acidic loop formed by residues 222–229 (refs 6 and 7). Leahy *et al.*



a, Comparison of the domains and membrane organization of CD4 and CD8. The immunoglobulin-variable-like domains for which crystal structures were obtained^{1–3} are in black. Although part of the glycosylated 'hinge' was part of the CD8 fragment crystallized, it was not sufficiently ordered to yield a structure¹. *b*, Comparison of the relative orientation of the pairs of immunoglobulin (Ig)-like domains in CD4 and CD8. The approximate positions of the three CDR-like loops in each Ig-like domain are shown. In the CD8 dimer (left) the placement of these domains is similar to those of the CDR loops of heavy and light chains of antibodies. By dislocation (arrow) one can convert the CD8 box into the CD4 rod. The proposed sites of interaction of CD8 with class I MHC molecules, and CD4 with class II MHC molecules and the HIV gp120 protein, are indicated by the brackets.

illustrate how this loop might dock with the CDR loops of CD8 which are predominantly positively charged. Uncertain as yet are the precise areas of contact or the stoichiometry of the interaction. The observation by Krishna *et al.*⁸ of "tetrameric cell-surface MHC class I molecules" further stimulates consideration of Leahy and colleagues' proposition that one CD8 α homodimer may bind two class I molecules.

At first glance, the structure of the CD4 fragment appears to be quite distinct from that of CD8. Although the folds of the individual Ig-variable-like domains are similar, the two domains of CD4 abut in a longitudinal fashion forming "a rod-shaped molecule"². The basis of this novel pairing of Ig-like domains is an extended β -strand, one end of which contributes the final strand of the first domain while the other provides the first strand of the second domain. On further examination, however, similarities emerge in the topology of the two domains in the CD4 and CD8 structures — simplis-

tically, one can envisage CD4 as a dislocated version of CD8 (or vice versa) in which a relative rotation of the two domains has allowed alignment and joining of the last strand of one domain with the first strand of the other (*b* in the figure).

A consequence of the rod-like structure is that the CDR-like regions of the two CD4 domains are not brought into register as in CD8. Despite their spatial separation, the CDR-like loops from both domains (CDR1 and 3 of domain 1, CDR3 of domain 2) contribute to the binding site for class II MHC molecules as revealed by the mutational analysis of Fleury *et al.*⁹. In contrast, the binding site for the HIV glycoprotein gp120 localizes to the CDR2-like loop of the

amino-terminal CD4 domain (*b* in the figure).

Thus, from mutagenesis of the coreceptors, the MHC-binding site of CD4 appears to be more extended than that of CD8, and different combinations of CDR-like loops may be used for interaction with the MHC in the two molecules. On the other hand, from the MHC side, the two sets of interactions have features in common. Two papers provide evidence that the CD4-binding

1. Leahy, R., Axel, R. & Henderickson, W. A. *Cell* **68**, 1145–1162 (1992).
2. Wang, J. *et al.* *Nature* **348**, 411–418 (1990).
3. Ryu, S.-E. *et al.* *Nature* **348**, 419–426 (1990).
4. Littman, D. R., Thomas, Y., Maddon, P., Chess, L. & Axel, R. *Cell* **240**, 237–246 (1985).
5. Sanders, S. K., Fox, R. O. & Kavathas, P. J. *exp. Med.* **174**, 371–379 (1991).
6. Salter, R. D. *et al.* *Nature* **345**, 41–46 (1990).
7. Connolly, J. M., Hansen, T. H., Ingold, A. L. & Potter, T. A. *Proc. natn. Acad. Sci. U.S.A.* **87**, 2137–2141 (1990).
8. Krishna, S., Benaroch, P. & Pillai, S. *Nature* **357**, 164–167 (1992).
9. Fleury, S. *et al.* *Cell* **66**, 1037–1049 (1991).
10. König, R., Huang, L. Y. & Germain, R. N. *Nature* **356**, 769–798 (1992).
11. Cammarota, G. *et al.* *Nature* **356**, 799–801 (1992).

site of class II MHC molecules is the homologous loop — residues 137–143 of the class II β -chain — to that of class I which interacts with CD8. König *et al.*¹⁰ establish this point by mutagenesis of the β -chain; Cammarota *et al.*¹¹, in a complementary study, show that peptides derived from this region bind to CD4. Exactly how these loops interact with

CD4 and CD8, and whether they provide the only sites of contact, are questions awaiting the cocrystallization of MHC molecules and their coreceptors. □

Peter Parham is in the Departments of Cell Biology, and Microbiology and Immunology, Stanford University School of Medicine, California 94305, USA.

GENE REGULATION

Ecdysone and the onion

Greg Guild and Geoff Richards

THE axiom that more than half of the secret of finding something is knowing where to look was well illustrated by two meetings* devoted to the molecular analysis of steroid response in insects. By exploiting the combined genetic and *in vivo* approaches available with *Drosophila*, levels of detail in the steroid regulation of gene networks are emerging that are rarely seen elsewhere.

Although they are structurally related to vertebrate steroids, the insect ecdysones seem to be harmless to mammals even at their relatively high physiological levels of 10^{-6} M. That stimulated many insect physiologists to characterize the steroid response with the aim of producing ecdysone derivatives which might prove to be specific regulators of insect populations. Their ingenuity in devising both *in vivo* and *in vitro* test systems using ligatures, microsurgery and organ culture has resulted in a detailed understanding of the ecdysone response. But the small size of insects proved to be a liability in the biochemical isolation of the ecdysone receptors, and the molecular breakthrough finally came from a long-term *Drosophila* project.

Molecular analysis of the ecdysone response in *D. melanogaster* was stimulated by the chromosome 'puffing' model of Ashburner *et al.*¹, and started with the isolation of genes from defined puffs of the giant polytene chromosomes. Following studies which revealed that three of the primary ecdysone responsive puffs encode different families of transcription factors^{2,3}, the ecdysone receptor itself has now been shown to be present in at least three isoforms (M. Koelle, W. Talbot and D. Hogness, Stanford Univ.). The combinatorial possibilities of isoforms of both the receptor and the primary responsive genes are much larger than those suggested by the formal model, and the present aim is to understand the stage- and tissue-specific response to a single hormonal stimulus (ecdysone). If the available molecular

complexity proves insufficient to explain that response, there are further potential players in the receptor-related molecules in *Drosophila* now under study (V. Heinrich, Univ. North Carolina, Greensboro), and in the new primary responsive genes that have been isolated (P. Hurban and A. Andres, Univ. Utah).

Comparative studies in other insects such as *Manduca sexta* (L. Riddiford, Univ. Washington, Seattle; W. Segraves, Salk Inst.) and *Galleria mellonella* (M. Jindra, Caske Budovice, Czechoslovakia) show that the *D. melanogaster* organization of these alternatively spliced genes is not unique. This bears on the interpretation of certain novel isoforms (for example one-fingered receptors)⁴ and opens the possibility of extending the results to insect species of medical or economic importance.

Every good nuclear receptor needs a DNA-binding site, and a clearer picture of the elusive EcRE (ecdysone response element) is now emerging. Although closely related to the model hormone response elements of vertebrates, the insect elements are not necessarily defined solely by consensus sequences (P. and L. Cherbas, Indiana Univ.; C. Antoniewski and M. Laval, Inst. Jacques Monod) and most probably require interactions between receptors and other transcription factors for their specific activity *in vivo*.

One striking feature of the meetings was the return to the *in vivo* analysis of mutants, necessarily left to one side during the difficult molecular analyses of the monster primary response genes whose isoform transcripts extend over 100 kilobases. Particularly notable was the report of the rapid molecular characterization of fly strains bearing receptor mutants induced by classical mutagenesis (M. Bender and D. Hogness). Equally, mutants in two of the early genes, the *Broad Complex* (J. Deutsch, Inst. Jacques Monod; G. G.; R. Hodgetts, Univ. Alberta; L. Von Kalm, Univ. California, Berkeley; F. Karim, Univ. Utah) and E74 (J. Fletcher, Univ. Utah) have been used to show their importance for the

transcription of a number of downstream genes.

Unsuspected detail in the *Drosophila* ecdysone response *in vivo* is becoming apparent from the use of techniques, such as immunocytology, with isoform-specific antibodies for ecdysone receptors (J. Truman, Univ. Washington, Seattle, with W. Talbot and D. Hogness), or transcript analyses using finely staged material by both classical northern analyses (F. Karim; C. Bayer, Univ. California, Berkeley) or simultaneous RTase-PCR analysis of isoforms of different primary response genes in tissues of individual *Drosophila* larvae (G.R.). These studies show that adjacent cells in the same tissue may express different receptor isoforms; that isoform switching in the primary response genes is a rapid, but sometimes transient, response to hormone; and that there are clear differences in the regulation of the primary response in different tissues of individuals. Similar developmental studies in a number of tissues have revealed an ecdysone response in the mid-third instar of *Drosophila*, which occurs before the dramatic hormone responses that initiate pupariation and metamorphosis. Molecular characterization of this response will be necessary to understand the reprogramming of cells for the later ecdysone responses. Products of the *Broad Complex* locus are also involved in many of the mid-third-instar responses — but not all of them, as shown for the ecdysone-dependent modulation of the dipterin promoter's response to bacterial infection (M. Meister, Univ. Louis Pasteur), a first tentative link between the endocrine and immune systems in insects.

For old hands all this showed that we have penetrated another layer of the onion and, as best expressed by Jim Truman, we are in for a period of intense data collection before we can hope to explain the remarkably rich biological complexity, both in time and space, that characterizes the regulation of these gene networks *in vivo*. It may well be that the vertebrate laboratories, which have dominated the last decade of steroid research, will have something to learn from these studies. □

Greg Guild is in the Department of Biology, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6018, USA. Geoff Richards is in the LGME du CNRS/U184 de l'INSERM/Institut de Chimie Biologique, 67085 Strasbourg Cedex, France.

* Molecular Biology of Ecdysone Response Philadelphia, 11 March, 1992. Tenth Ecdysone Workshop, University of Liverpool, 6–9 April, 1992.

1. Ashburner, M. *et al.* Cold Spring Harb. Symp. quant. Biol. **38**, 655–662 (1974).
2. Thummel, C. S. *BioEssays* **12**, 561–568 (1990).
3. Andres, A. & Thummel, C. S. *Trends Genet.* **8**, 132–138 (1992).
4. Segraves, W. A. *Cell* **67**, 225–228 (1991).

Looking inside quantum dots

L. Eaves

MINIATURIZATION in microelectronics has reached a point where researchers are trying to create quantum dots, more fancifully dubbed artificial atoms, embedded in semiconductor structures. The energy levels of electrons in these dots, as in real atoms, are discrete not continuous, but can be varied at will by the experimenter. R. C. Ashoori *et al.* now show (*Phys. Rev. Lett.* **68**, 3088–3091; 1992) that it is possible to measure the spectrum of a quantum dot by recording the small current that flows as a single electron enters or leaves it.

The techniques that make this research possible are the techniques that have given us key components in two of the most popular products in the consumer electronics market — compact-disk players and satellite-television receiver dishes. The essential feature of the components (quantum-well lasers and high-electron-mobility transistors) is the two-dimensional nature of their electronic states. Confinement provided by planar potential-energy barriers in the semiconductor structure constrains electrons to move freely in two dimensions only. In the third direction, perpendicular to the barriers, the motion is fully quantized — one-half of a de Broglie wavelength fits into the narrow quantum well (perhaps 100 Å across) separating the two barriers. This lowering of the dimensionality of the electron motion

with the prospect that the level separation can be modified in a controllable way.

The quantum dot of Ashoori *et al.* is based on a two-dimensionally confined electron gas, formed in a thin (15 nm) quantum-well layer of gallium arsenide (GaAs) sandwiched between two layers of aluminium gallium arsenide (AlGaAs). The additional confinement is provided by applying a voltage to a chromium-metal gate electrode deposited on the surface of the semiconductor heterostructure. When this is biased negative, the electronic motion in the quantum well is restricted to a disk-shaped quantum dot with a diameter of less than 1 μm (Fig. 1). As the gate voltage is altered, electrons can enter the dot from a high-conductivity n⁺GaAs layer by quantum-mechanical tunnelling through the lower barrier, which is only 85 Å across.

The tunnelling occurs only when the characteristic 'Fermi' energy of the electrons in the lower GaAs layer equals one of the distinct energy levels of the quantum dot. This allows Ashoori *et al.* to measure the 'spectrum' of their artificial atom. They start their experiment with the quantum dot empty. By steadily decreasing the gate voltage, V_G , they can add electrons one-by-one to the quantum dot. The particular voltages, V_G , at which this occurs correspond to

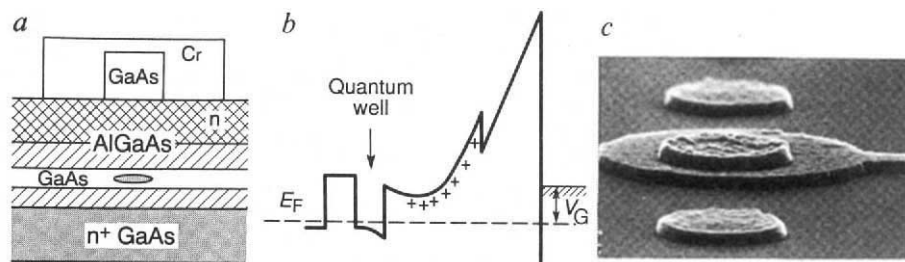


FIG. 1 *a*, A schematic cross-section through Ashoori *et al.*'s heterostructure, sandwiched by a chromium and a high-conductivity (n⁺GaAs) semiconductor electrode. The quantum dot (shaded ellipse) is confined in the GaAs quantum well. *b*, Potential energy diagram showing the potential well provided by the sandwiched GaAs layer, and the Fermi energy (E_F). *c*, Micrograph (2.5 μm across) of the device, showing the top chromium gate.

gives rise, in part, to the components' special properties.

Confinement allows physicists and electronic engineers to reduce the dimensionality further, to create the quantum dot in which the electrons' motion is restricted and quantized along all three directions. In this case the energy states of the electrons in the semiconductor should occupy discrete levels rather than the continuous energy bands which are characteristic of bulk solids. In other words, quantum dots should have atom-like energy levels, but

the internal energy levels of the dot. The actual measurements involve an a.c. capacitance technique in which an electron can be moved back and forth between the quantum dot and the GaAs electrode.

Figure 2 shows a typical experimental sweep. A sequence of well resolved and reproducible peaks in the capacitance is observed. Each of the smaller peaks (such as those at gate voltages around -600 mV) arises from a single electron tunnelling into one of the low-energy discrete energy levels of the quantum

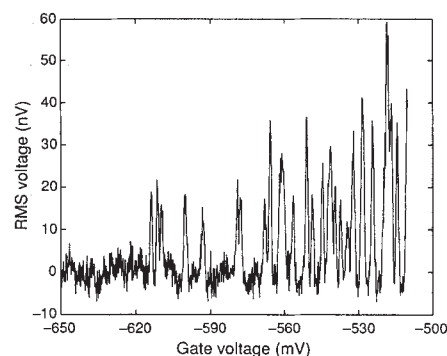


FIG. 2 The root-mean-square signal as a function of gate bias using 0.33-mV a.c. excitation. A linear background has been subtracted.

well. The trace therefore represents a direct energy-level spectrum of the quantum dot. The spacing of the peaks in Fig. 2 does not correspond exactly to the spacing of the energy levels, owing to an 'electrostatic leverage' effect.

But the characteristic level spacing is a few millielectronvolts. This is too large a value to arise from the quantum confinement provided by the electrostatic potential of the gate itself.

Instead, it is thought that potential-energy fluctuations in the quantum well create 'puddles' which further localize the electrons. This interpretation is supported by investigating the variation of the peak position (V_G) when a large magnetic field is applied perpendicular to the plane of the two-dimensional electron gas. The effect of the field is to provide even stronger confinement on the energy puddles. By looking at how the spectrum varies with applied field, the authors identify two scales for the potential puddles, both attributable, they suggest, to silicon atoms used to dope the GaAs electrode. Weak, shallow puddles, responsible for confinement energies of 0.3–2.5 meV, are probably the result of variations in silicon concentration in or near the GaAs electrode.

Stronger confinement, indicated by a weak dependence on applied field, is probably due to silicon atoms that migrated from the electrode into the well as the device was fabricated. Indeed, my colleagues have found that individual impurity atoms in quantum dots can completely dominate many of the dot's characteristics (M. W. Dellow *et al.* *Phys. Rev. Lett.* **68**, 1754–1759; 1992; see S. Washburn's News and Views, *Nature* **357**, 199–200; 1992; and also P. Guéret *et al.* *Phys. Rev. Lett.* **68**, 1896–1899; 1992 for a contrary view). The ultimate, unadulterated artificial atom is yet to be created. □

L. Eaves is in the Department of Physics, University of Nottingham, Nottingham NG7 2RD, UK.

Z and the insoluble answer

Richard N. Sifers

ACCUMULATION in the liver of the Z variant of human α_1 -antitrypsin can lead to cirrhosis, but the structural identity of the accumulated protein has been unknown. On page 605 of this issue¹, Lomas *et al.* unveil a totally unexpected finding — the Z variant forms insoluble homopolymers in the endoplasmic reticulum of the liver, and does so by a well-defined mechanism.

Human α_1 -antitrypsin (AAT) is a single, folded polypeptide that inhibits the action of serine proteases. Although the macromolecule is one of the main components of serum, its predominant physiological role is to prevent degradation of elastin fibres in the lung. It accomplishes this task by forming a tight complex with elastase, a protease secreted from neutrophils in the lung, which ultimately inhibits its hydrolytic activity². Heritable forms of AAT deficiency are most often associated with the synthesis of a variant macromolecule, which bears a mutation in the polypeptide³ that somehow hinders its ultimate secretion from hepatocytes⁴ (the principal site of synthesis). Decreased secretion lowers the level of serum AAT, and also the total elastase inhibitory activity in the lung. The resulting degradation of elastin fibres is implicated in emphysema².

A subset of AAT variants accumulates within the endoplasmic reticulum of the liver, and is associated with cirrhosis⁵. One such variant, designated Z, contains a glutamate-to-lysine substitution at residue 342 (ref. 2). Because of its frequency in the population, and the severity of its impaired secretion, the Z variant has served as a prototype for analysing intrahepatic protein retention and accumulation.

Discovery of the Z-variant accumulation mechanism resulted from analysis of a structural feature shared by this and other members of a class of serine protease inhibitors known as serpins⁶. Briefly, the reactive inhibitory site of serpins centres on two amino acids in a 14-amino-acid loop at the surface of the folded polypeptide^{6,7}. For AAT and other inhibitory members of this family, amino-terminal peptides of the reactive centre loop are inserted into a gap in the major structural feature of the folded macromolecule, β -pleated sheet A (see Fig. 2 of the paper of Lomas *et al.*, page 606). This 'stressed' conformation is required for the reactive centre to function as a substrate to react with and inhibit targeted proteases. Indeed, the absence of inhibitory activity for some members of the serpin family may result from the substitution of amino acids within the

reactive centre loop, thereby preventing its partial insertion into sheet A and the formation of the stressed conformation⁷.

A clue to the mechanism of Z accumulation in liver was reported by Carrell *et al.*⁸ last year, following their observation that under mild denaturing conditions the entire reactive centre-loop peptide of an intact inhibitor, such as AAT, could 'lock' into sheet A. This event is akin to the major structural rearrangement of AAT when the loop has been relaxed by proteolytic cleavage at its reactive inhibitory site and actually becomes strand 4 of the sheet. As described in the new report¹, incubation of the purified, secreted form of the normal M variant under these conditions led to its unexpected polymerization *in vitro*. Polymerization between macromolecules pre-bound with a synthetic peptide homologous to the reactive centre loop, which had inserted into sheet A, did not occur. This and subsequent analyses implied that polymerization under these conditions results from the stable insertion of the mobile reactive centre loop of one molecule into sheet A of another. Increased polymerization of the Z variant (that is, spontaneous polymerization under non-denaturing conditions) was then observed *in vitro*, as the authors had predicted. This prediction was based on the ability of the glutamate-to-lysine substitution at the hinge region of the reactive centre loop to prevent its insertion into sheet A, thereby leaving it

in an 'opened' conformation ready to accept the appropriate loop from another Z molecule.

Confirmation that loop-sheep polymerization of the Z variant does occur in liver endoplasmic reticulum comes from the striking electron micrographs showing tangled polymers of the macromolecule purified from human liver biopsies (see Fig. 3, page 606). Furthermore, the morphology and physical characteristics of the liver-derived Z polymers are identical to those of the serum-derived protein polymerized *in vitro*. The authors did not test the effect of nonspecific peptides on *in vitro* polymerization of either the M or Z variant, but their hypothesis of loop-sheet polymerization is borne out by physical and biochemical analyses.

The tendency of the secreted Z variant to undergo spontaneous aggregation⁹ led most, if not all, researchers in the field to conclude that a similar mechanism was responsible for its retention and accumulation in the liver endoplasmic reticulum. But this conclusion was never quite secure because it could not explain why the macromolecule isolated from liver inclusion bodies could exhibit full protease inhibition¹⁰. Conceivably, a randomly aggregated molecule would be grossly misfolded and therefore display little, if any, inhibitory activity after isolation. The evidence for the accumulation of polymerized Z protein now satisfies this discrepancy.

In terms of protein retention in the endoplasmic reticulum, it is reasonable to conclude that polymerization of the Z variant could dramatically hinder its export from that compartment, as the

Out of the frying pan

THE Saharan silver ant *Cataglyphis bombycina* treads a narrow line between heat exhaustion and being eaten by lizards. As R. Wehner *et al.* demonstrate on page 586 of this issue, the ants prefer to run the risk of the former than face the near-certainty of the latter. Most desert creatures, lizards included, seek shade when surface temperature exceed about 45 °C. In response, the ants emerge to forage only when temperatures rise above 46.5 °C, and can tolerate temperatures of up to about 53.6 °C. Confined to this narrow thermal 'window', they operate at the limits of their tolerance and adopt additional strategies to minimize the heat burden. The ant pictured here, for example, is resting on a stalk of grass where the ambient temperature is slightly less intense than that at ground level.



H. G.

authors suggest. But one must now consider a new question — quite simply, is loop-sheet polymerization used for retaining all secretion-impaired AAT variants or is there an alternative mechanism in some cases? Because most secretion-impaired variants do not show any evidence of accumulating in hepatocytes^{3,4}, the second hypothesis has to be valid. For example, how else can one explain the retention mechanism employed for the truncated Null^{Hong Kong} variant, in which the entire reactive centre loop is missing¹¹?

Overall, given that molecular chaperones are increasingly implicated in catalysing the folding and causing the retention of a variety of secretory and membrane-associated polypeptides in the liver endoplasmic reticulum¹², it might be premature to imply that loop-sheet polymerization is the sole mechanism for retaining the Z variant, or any other AAT variant. The identification of a soluble, high-molecular-weight form of the newly synthesized Z variant in transfected mouse hepatoma cells¹³ might reflect its association with a chaperone, rather than support the hypothesis suggested by Lomas *et al.*¹.

What, if anything, does the discovery of polymerized Z protein tell us about AAT-related liver disease? Apparently, accumulation of the insoluble human Z variant in hepatocytes of transgenic mice is enough to induce liver damage in those animals¹⁴. Unexpectedly, the normal M variant of the human protein can also accumulate in the hepatic endoplasmic reticulum in transgenic mice, but only as a result of its increased synthesis¹⁵. Unlike the Z variant, the accumulated M protein is completely soluble following cell lysis¹⁵ and these animals exhibit no apparent liver damage¹⁴. This observation raises the

possibility that it is the insolubility of the accumulated Z variant that might actually be the cause of liver damage, at least in mice. However, understanding the mechanism of Z accumulation should at least aid in the design of methods to suppress that event, and allow all the retained protein to be degraded¹³.

As a possible way forward from Lomas and colleagues' findings, it has been observed that accumulation of the normal M variant in the liver endoplasmic reticulum causes a concomitant accumulation of the endogenous murine AAT polypeptide¹⁵. Although this might represent the formation of human-

mouse AAT heteropolymers, the exact identity of this material remains to be determined. But these findings do show that transgenic mice might be a helpful model for further analysing the loop-sheet polymerization mechanism as it pertains to the association of nonidentical serpin molecules in the endoplasmic reticulum, and also to test methods for suppressing the accumulation of the polymerized Z variant. □

Richard N. Sifers is in the Departments of Pathology and Cell Biology, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA.

SOLAR SYSTEM

Wandering on a leash

Carl D. Murray

MENTION of chaos in the motion of objects in the Solar System conjures up images of colliding planets, impacting asteroids and a variety of catastrophic events. However, a number of recent discoveries have shown that although gravitational perturbations can lead to chaotic orbits, the nature of the chaos can be quite subtle. A numerical investigation of the orbit of the asteroid 522 Helga, reported by Milani and Nobili on page 569 of this issue¹, shows that it is probably an example of 'stable chaos', where the orbit is chaotic yet confined to a particular region of the asteroid belt. Although the concept may seem counter-intuitive, stable chaotic motion has now been identified in a number of orbits in the Solar System.

The explanation of this apparent paradox has more to do with the imprecision of language than any fundamental problem with the dynamics. Using everyday definitions it is difficult to envisage a political or economic situation that is simultaneously stable and chaotic; they appear to be mutually exclusive states. However, using more precise definitions, such a combination is commonplace in a number of dynamical systems. Chaos can be thought of as a sensitive dependence on initial conditions, or the exponential divergence of nearby trajectories and the rate at which this occurs can be calculated. A definition of 'stable' is more difficult because it means different things to different people in different circumstances. For example, an incomplete list by Szebehely² gives 47 different concepts of stability. An intuitive definition is that a given system is stable if it does not deviate from its original configuration by more than some specified amount. In this sense a stable orbit is one which is confined or bounded such that deviations take place within strict limits.

Stable chaos occurs when such deviations are chaotic. A simple analogy is the motion of a ball on a roulette wheel; although dissipation and the skill of the croupier are important, the ball's trajectory is chaotic yet usually confined to the vicinity of the wheel.

Examples of stable chaos can be seen in a variety of dynamical systems ranging from simple systems with one degree of freedom to the highly complex gravitational interactions of the planets. Numerical investigations of the orbit of the outer planets^{3,4} by Wisdom and colleagues showed good evidence that the orbit of Pluto is chaotic with a timescale for exponential divergence of 20 million years (Myr). Laskar^{5,6} used an integration of the averaged equations of motion to show that the divergence timescale for the inner planets is as low as 5 Myr. In neither case was there any indication of gross instability; the planets remain in their orbits and the chaos just acts to limit the predictive power of celestial mechanics. However, conclusions from such numerical integrations are only as good as the results for the last time step — as yet there is no numerical or analytical proof that the orbits of the planets are stable for timescales approaching the age of the Solar System, 5 billion years.

522 Helga is not the first object in the Solar System found to be exhibiting stable chaos, but the result of Milani and Nobili is unusual in two respects: the rate of orbit divergence and the location of the stable chaos. Although the integration time was short (7 Myr), it corresponds to 1,000 times the divergence timescale, compared with factors of 50 and 40 obtained for Pluto and the inner planets respectively. This gives the authors the confidence to state that the orbit of the asteroid is an example of stable chaos. But it has to be remem-

1. Lomas, D. A., Evans, D. L., Finch, J. T. & Carrell, R. W. *Nature* **357**, 605–607 (1992).
2. Gadek, J. E. & Crystal, R. G. *Metabolic Basis of Inherited Disease*, 5th edn (McGraw-Hill, Minneapolis, 1982).
3. Brantly, M., Nukiwa, Y. & Crystal, R. G. *Am. J. Med.* **84**, 13–31 (1988).
4. Sifers, R. N., Finegold, M. J. & Woo, S. L. C. *Am. J. Respir. Cell molec. Biol.* **1**, 341–345 (1989).
5. Sharp, H. L., Bridges, R. A., Krivit, W. & Freier, E. F. *J. Lab. clin. Med.* **73**, 934–939 (1969).
6. Huber, R. & Carrell, R. W. *Biochemistry* **28**, 8951–8966 (1989).
7. Wright, H. T., Quian, H. X. & Huber, R. *J. molec. Biol.* **213**, 518–528 (1990).
8. Carrell, R. W., Evans, D. L. & Stein, P. E. *Nature* **353**, 576–578 (1991).
9. Cox, D. W., Billingsley, G. D. & Callahan, J. W. *FEBS Lett.* **205**, 255–260 (1986).
10. Bathurst, I. C., George, P. M., Travis, J. & Carrell, R. W. *FEBS Lett.* **117**, 179–183 (1984).
11. Sifers, R. N., Brashears-Macatee, S., Kidd, V. J., Muensch, H. & Woo, S. L. C. *J. biol. Chem.* **263**, 7330–7335 (1988).
12. Gething, M. J. & Sambrook, J. *Nature* **355**, 33–44 (1992).
13. Le, A., Ferrell, G. A., Dishon, D. S., Le, Q.-Q. & Sifers, R. N. *J. biol. Chem.* **267**, 1072–1080 (1992).
14. Carlson, J. A. *et al. J. clin. Invest.* **83**, 1183–1190 (1989).
15. Sifers, R. N., Rogers, B. B., Hawkins, H. K., Finegold, M. J. & Woo, S. L. C. *J. biol. Chem.* **264**, 15696–15700 (1989).

bered that they integrated for less than 0.2 per cent of the age of the Solar System and their result is only a good indication of Helga's long-term stability rather than proof of it.

Jupiter's gravitational pull is responsible for the creation of a number of large, chaotic regions in the asteroid belt, primarily at orbital radii where the period is a simple fraction of Jupiter's period. Large changes in the orbits can occur at such radii and at 2.5 astronomical units (nearly half the distance of Jupiter's orbit), asteroidal material in chaotic zones can be injected into high eccentricity orbits that cross the Earth's path to become meteorites⁷. Therefore chaos in the asteroid belt is usually associated with an absence of material. However, the orbital periods of 522 Helga and Jupiter are in an approximate 7:12 ratio and a peculiarity of the resulting resonance mechanism acts to protect the asteroid from close approaches to Jupiter¹.

All the above examples of stable chaos in the Solar System are associated with resonance. Pluto is in 2:3 resonance with Neptune as well as a variety of more intricate relationships³, and the planets of the inner Solar System are involved in a number of long-period or secular resonances⁶. The association of stable chaos with such regions is not too surprising as the boundaries of resonance are known to be sites of chaotic motion (stable and otherwise) and this is observed even in the simplest dynamical systems. Regions where adjacent resonances overlap are also likely places to find chaos.

Undoubtedly there are many more examples of chaotic motion in the Solar System to be found and there will be many more numerical integrations to search for them. This branch of celestial mechanics has almost become an experimental science with the analytical results lagging far behind, but perhaps this is to be expected from a problem which lends itself so easily to numerical solutions. Nevertheless this is an area of research which could have an important impact closer to home; the Earth may not undergo large variations in its orbital elements, but it would be comforting to have similar guarantees for all the minor bodies of the Solar System. □

Carl D. Murray is in the Astronomy Unit, Queen Mary and Westfield College, University of London, Mile End Road, London E1 4NS, UK.

One thousand families for the molecular biologist

Cyrus Chothia

How many families of proteins are there? By putting together the information to be found in papers published over the past few months we can make an initial estimate, and my calculation suggests that the large majority of proteins come from no more than one thousand families.

Proteins are clustered into families whose members have diverged from a common ancestor and so have similar folds; they also usually have similar sequences and functions. This classification is central to our understanding of how life has evolved, and makes the

this work and the proportion of them that have sequences similar to those determined previously.

The six sets of results give a remarkably consistent view: in each case the authors report that close to one-third of the new sequences belong to families that already have members in the sequence databanks.

The number of protein families represented in the sequence databanks can be calculated approximately from the results reported by Sander and Schneider⁶ and by Pascarella and Argos⁷. Many of the protein structures that have been

TABLE 1 New gene sequences that are related to previously determined sequences

Genome projects			
Source	Total number of genes	Genes related to those previously determined	Ref.
<i>Caenorhabditis elegans</i> chromosome III (part)	32	14 (44%)	1
Yeast chromosome III	182	52–66 (29–36%)	2
chromosome IX (part)	46	15 (33%)	*
Large libraries of expressed genes			
Source	Total number of clones	Clones related to previously determined protein sequences	
Human brain	~1,400†	406 (~30%)	3
<i>Caenorhabditis elegans</i> St Louis–Cambridge	1,517	512 (34%)	4
NIH	585	210 (36%)	5

* B. Barrell, V. Smith and C. Brown, personal communication.

† Adams *et al.*³ sequenced 2,375 cDNA clones, of which 406 are homologous to previously known sequences. They estimate that not more than half of the remainder contain coding sequences.

elucidation and definition of such families one of the principal concerns of molecular biology. With the publication of the first results of the yeast, nematode and human genome projects^{1–5}, and the availability of two new analyses of known protein structures^{6,7}, we can begin to put that endeavour on a numerical footing.

The recent reports have described the genes present in the whole of yeast chromosome III and part of chromosome III of the nematode *Caenorhabditis elegans*^{1,2}. The genes in part of yeast chromosome IX have been determined by B. Barrell, V. Smith and C. Brown (personal communication). In addition to this genome work, large libraries of human and *C. elegans* expressed genes have been sequenced^{3–5}. Table 1 shows the total number of genes identified by

determined by X-ray crystallography and NMR are stored in the Brookhaven databank, and Sander and Schneider⁶ determined the proportion of the entries in the EMBL/SwissProt sequence databank that are homologous to the sequences of proteins stored at Brookhaven. This work was originally carried out on 1989 versions of the two databanks⁶ but Sander and Schneider have repeated their calculations using the current versions (personal communication). They find that 28% of the entries in the sequence databank have at least a 25% residue identity with one of the entries in the structure bank. This means that at least a quarter of the currently known protein sequences belong to protein families for which there are structures in the Brookhaven databank.

The number of protein families repre-

1. Milani, A. & Nobili, A. M. *Nature* **357**, 569–571 (1992).
2. Szebehely, V. *Celest. Mech.* **34**, 49–64 (1984).
3. Sussman, G. J. & Wisdom, J. *Science* **241**, 433 (1989).
4. Wisdom, J. & Holman, M. *Astr. J.* **102**, 1528–1538 (1991).
5. Laskar, J. *Nature* **338**, 237–238 (1989).
6. Laskar, J. *Icarus* **88**, 266–291 (1990).
7. Wisdom, J. *Nature* **315**, 731–733 (1985).

sented in the Brookhaven databank can be found from the results of Pascarella and Argos⁷, who grouped together the structures that share the same fold (that is, have the same secondary structures in the same orientation with the same chain topology). The 254 different proteins in the April 1991 issue of the Brookhaven structure databank have 83 different folds⁷. Now, proteins with the same fold need not be related: they could have the same fold because of the rules that govern secondary structure packings and chain topologies⁸. But in the large majority of the 83 groups the proteins also have sequence, functional and/or

TABLE 2 Genome projects

Species	Approximate number of genes	Tentative date of completion
<i>Escherichia coli</i>	4,000	1995–98
Yeast	7,000	2000
<i>Caenorhabditis elegans</i>	15,000	2000
Human	50–100,000	2015

genetic similarities that strongly imply that they are descended from a common ancestor. Consideration of the few instances where this is not the case, and of the many additions made over the past year, shows that the Brookhaven databank contains representatives of 120 different protein families.

So, to summarize, about a third of the sequences present in genomes are related to entries in the current sequence databank, and about a quarter of the current sequences belong to one of 120 protein families. Assuming that there is no strong bias in the databanks, and that the sequence comparisons give an accurate picture of family membership, this would indicate that there are some 1,500 different protein families. The fairly constant proportions of new sequences that correspond to old sequences (Table 1) suggest that the databanks are not strongly biased. However the sequence comparisons used in this work will underestimate the extent to which proteins are related.

Crystallography has made clear that proteins can evolve; so, though they continue to share the same fold, their residue identities are no greater than that of two randomly selected sequences. For example, actin has a sequence that has been strongly conserved in all eukaryotes and apparently had no relatives. Flaherty *et al.*⁹, however, show crystallographically that structural and functional similarities of actin and the ATPase fragment of the heat-shock cognate protein (forms of which are found in both prokaryotes and eukaryotes) are so close there can be little doubt that

they are descended from a common ancestor. The two proteins contain 375 and 386 residues, respectively, of which only 39 are identical.

Simple sequence comparisons with reasonable thresholds significantly underestimate the extent to which proteins are related (see also ref. 6). The calculation of the number of protein families described here is rather sensitive to such errors; for example, if we assume that the sequence comparisons find 80% of related proteins, and adjust the calculation accordingly, the number of protein families drops to 1,000. From current crystallographic results it seems that sequence comparisons are not this efficient. Thus a conservative view of the evidence that we have at present would be that the large majority of proteins come from no more than a thousand different families.

Extant proteins have been produced from the original set not just by point mutations, insertions and deletions but also by combinations of genes to give chimaeric proteins. This is particularly true of the very large proteins produced in the recent stages of evolution. Many of these are built of different combinations of protein domains that have been selected from a relatively small repertoire. The evolutionary mechanisms involved in their creation have been described by Patthy¹⁰.

In 1991 some 120 new protein structures were determined; in the previous year the number was 85. More than half of these were unrelated, or only very distantly related, to any previously known structure. Soon there will be 200 new structures each year. In Table 2, I list the current genome sequencing projects and the tentative dates of their completion. If most proteins do come from no more than one thousand families, crystallography, NMR and molecular modelling will produce, at least in outline, structures for most proteins in time for the completion of the genome projects. □

Cyrus Chothia is at the MRC Cambridge Centre for Protein Engineering and the Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK.

1. Sulston, J. *et al.* *Nature* **356**, 37–41 (1992).
2. Oliver, S. G. *et al.* *Nature* **357**, 38–46 (1992).
3. Adams, M. D. *et al.* *Nature* **355**, 632–634 (1992).
4. Waterston, R. *et al.* *Nature Genet.* **1**, 114–122 (1992).
5. McCombie, W. R. *et al.* *Nature Genet.* **1**, 124–130 (1992).
6. Sander, C. & Schneider, R. *Proteins* **9**, 56–68 (1991).
7. Pascarella, S. & Argos, P. *Prot. Eng.* **5**, 121–137 (1992).
8. Chothia, C. & Finkelstein, A. V. *A. Rev. Biochem.* **59**, 1007–1039. (1990).
9. Flaherty, K. M., McKay, D. B., Kabsch, W. & Holmes, K. C. *Proc. Natn. Acad. Sci. U.S.A.* **88**, 5041–5045 (1991).
10. Patthy, L. *Curr. Opin. struct. Biol.* **1**, 351–361 (1991).

Pure tones

LAST week Daedalus deduced that a molecule must be instantly evaporated by a photon which exactly supplies its latent heat of evaporation. He calculates that, for water at room temperature, photons of 2.71 micrometres would do the trick. For acetone, the resonant wavelength should be 3.66 micrometres, and for alcohol 3.04 micrometres. So, says Daedalus, shine an infrared laser tuned to 3.04 micrometres on a mixture of these three liquids, and only the alcohol will vaporize.

'Resonant distillation' will be a wonderfully direct and elegant method of chemical separation. Simply by stepping an irradiation laser through a set of frequencies, it will evaporate the components from the most complex mixture one at a time. Daedalus hoped to revolutionize the petrochemical industry with it, until he realized that no infrared laser is powerful enough. But for the bench-scale fractionation of liquid mixtures, resonant distillation should be unrivalled.

A properly tuned laser could quickly flash the traces of chloroform from tap water, or benzene from mineral water. It could neatly vaporize dioxin, caffeine, monosodium glutamate or other additives in foodstuffs. It could even lift the alcohol out of wine, while not disturbing its bouquet or the other subtle flavour-components which alone endear this regrettably intoxicating beverage to its connoisseurs.

More cunning still, resonant distillation is an analytical technique. If the liquid mixture is irradiated by audio-modulated infrared, the selected component will evaporate in a rapid sequence of pulses, giving an audible note whose intensity will reveal the concentration of that component. Tune the modulated laser through a range of wavelengths, and each of the volatiles present will sing out when its resonant wavelength is reached. A musical analyst might even irradiate a mixture with a battery of lasers, each tuned to a specific infrared wavelength and modulated by a particular note, and identify its components from the resulting musical chord. Unwanted components could then be 'sung to exhaustion' — irradiated till their answering note died into silence, showing that they had been completely evaporated.

This neat trick still works even if you don't know what the component is. Every chemical product has to be purified from unknown impurities. Once their frequencies have been identified, they could be sung to exhaustion. Only the desired product would remain, singing the strong single note that guarantees its purity.

David Jones

Tumours and Coley's toxins

SIR — Starnes¹ discusses the treatment of tumours with Coley's toxins, and makes the interesting point that 90% of Coley's successes occurred in patients with tumours of mesodermal origin, perhaps attributable to a greater probability that such tumours are immunogenic. Then, as there is some evidence that tumour necrosis factor (TNF) is also more effective against sarcomas, Starnes assumes a strong parallel between TNF therapy, endotoxin therapy and the use of Coley's toxins. But any therapy relying on the immune response is more likely to be effective if the tumour is immunogenic, so this argument does not indicate that Coley's toxins worked via TNF; indeed, there are reasons for suggesting that the mechanism was more complex.

Although TNF can be directly toxic for tumour cells, this is a rare mode of action *in vivo*, where the antitumour effects seem to operate via the tumour vasculature². Thus the functional state of the endothelium, or perhaps a minor degree of pre-existing inflammation in some tumours, results in a TNF-sensitive state analogous to that induced in the "prepared sites" described by Schwartzman³. Schwartzman injected bacterial suspensions into the skin of rabbits, and 24 h later injected a similar endotoxin-rich suspension intravenously. The result, after a further 20–24 h, was haemorrhagic necrosis confined to the previously injected skin site.

This response was almost certainly mediated by systemically released TNF reaching the "prepared site", as similar necrosis can be evoked in such sites by the direct injection of TNF^{4,5}. When a bolus of the cytokine reaches such a site there is rapid destruction of the vasculature, and subsequently of the dependent tissue. Infusion of large doses of TNF into perfused limbs together with interferon- γ and melphalan has reproduced this effect in human malignant melanomata and sarcomata⁶. In animals, this phenomenon, like the Schwartzman reaction in sites 'prepared' with lipopolysaccharide (LPS) or endotoxin, or other bacterial extracts, is rapid. The necrosis is seen about 24 h after TNF (or TNF-releasing trigger) is given. Moreover, after sublethal doses of LPS or TNF in several experimental systems, TNF-mediated tissue damage cannot be elicited again for at least 24 h (refs 5, 7, 8). The mechanism of this refractory state may be tachyphylaxis (though there is disagreement as to whether this affects antitumour activity^{9,10}), temporary exhaustion of the TNF-releasing potential of macrophages, or the triggering of counteracting physiological pathways

such as cortisol secretion and free TNF receptors.

Does this fit the descriptions of Coley's observations¹¹? First, Coley's effects were usually seen after 2–3 weeks, not after 24 h, and in some cases no tumour regression was seen until the toxin administration had continued for several months (for example, the case described on page 84 of ref. 11). Second, the toxins were usually administered daily or on alternate days, so for these reasons a 'Schwartzman-like' mechanism is improbable. Finally, before *Serratia marcescens* (which contains LPS and other toxic entities) was added to the formulation, results were seen by causing a skin infection with group A streptococci (erysipelas), or following injection of the group A *Streptococcus* alone. Is there something special about this organism in this context? In the pre-antibiotic era, when infection of ulcerating tumours was common, was remission also observed following infection with LPS-containing Gram-negative bacteria (which must have been common), or only if the organism was a group A *Streptococcus*?

Protein crystallization cont . . .

SIR — The discussion by Pitts^{1,2} and Abad-Zapatero³ on the crystallization of proteins by centrifugation deserves some further comments, if only to clarify some misunderstandings. The crystallization by centrifugation reported by Wyckoff and Corey⁴ in 1936 was not of tobacco mosaic virus (TMV) coat protein but was a pellet of the ribonucleoprotein viral rods with X-ray diffraction properties similar to those of the liquid crystals or tactoids of TMV previously observed by Stanley⁵. TMV solutions/suspensions at concentrations above 50 mg per ml under low salt conditions will separate at 1g into two phases; the lower, concentrated phase exhibits some properties of liquid crystals (for a description of the behaviour of concentrated solutions of TMV see refs 6, 7). Centrifugation only enhances this separation, but true three-dimensional crystals of TMV rods are not obtained. On the other hand, crystals of TMV coat proteins suitable for

I agree strongly with Starnes' speculation that there is more to be learnt about Coley's fascinating observations, and I suggest that we should focus attention on the group A *Streptococcus* itself, and on novel effects of chronic release of many cytokines, rather than on short-term effects of single large doses of TNF.

Graham Rook

Department of Medical Microbiology,
University of Middlesex

School of Medicine,
67-73 Riding House St,
London W1P 7PP, UK

1. Starnes, C.O. *Nature* **357**, 11–12 (1992).
2. Havell, E. A., Fiers, W. & North, R. J. *J. exp. Med.* **167**, 1067–1085 (1988).
3. Schwartzman, G. *Phenomenon of Local Tissue Reactivity and its Immunological Pathological and Clinical Significance*, 1–461 (Hoeber, New York, 1937).
4. Rothstein, J. & Schreiber, H. *Proc. natn. Acad. Sci. U.S.A.* **85**, 607–611 (1988).
5. Al-Attiah, R., Rosen, H. & Rook, G. A. W. *Clin. exp. Immun.* (in the press).
6. Lienard, D., Ewalenko, P., Delmotte, J. J., Renard, N. & Lejeune, F. J. *J. clin. Oncol.* **10**, 52–60 (1992).
7. Freudenberg, M. A. & Galanos, C. *Infect. Immun.* **56**, 1352–1357 (1988).
8. Sheppard, B. C., Fraker, D. & Norton, J. A. *Surgery* **106**, 156–161 (1989).
9. Takahashi, N., Brouckaert, P. & Fiers, W. *Cancer Res.* **51**, 2366–2372 (1991).
10. Fraker, D. L., Sheppard, B. C. & Norton, J. A. *Cancer Res.* **50**, 2261–2267 (1990).
11. Nauts, H. C., Fowler, G. A. & Bokatkot, F. H. *Acta med. scand. Suppl.* **274–277**, 1–93 (1953).

X-ray diffraction analysis were obtained by conventional means from ammonium sulphate solutions much later⁸.

The salient point about Pitts' procedure¹ is that ultrafiltration by centrifugation led to a condition of supersaturation for his protein under the particular solvent conditions. This is, in principle, no different from the results obtained with *Penicillium vitale* catalase in a solution containing 2-methyl-2,4-pentandiol⁹. The effectiveness of a protein precipitant is, generally, a linear function of the concentration of the precipitant and the logarithm of the solubility of the protein in a relationship that resembles the Setschenow equation for salting-out¹⁰. In the case of *P. vitale* catalase⁹, the initial concentration of the protein was very low and because of its relatively low molecular mass, a long centrifugation at high *g* forces would be required to establish a sufficiently high concentration of protein for supersaturation under the particular solvent condition. One must recognize the individuality of proteins; their maximum solubilities under various solvent conditions are part of this individuality. Thus it is unlikely that, in either case, crystals of these proteins have been obtained by centrifugation *per se* but rather that centrifugation produced a critical protein concentration favouring crystallization.

The important point of the observa-

1. Pitts, J. E. *Nature* **355**, 117 (1992).
2. Pitts, J. E. *Nature* **356**, 392 (1992).
3. Abad-Zapatero, C. *Nature* **356**, 392 (1992).
4. Wyckoff, R. W. G. & Corey, R. B. *Science* **84**, 513 (1936).
5. Stanley, W. M. *Science* **81**, 644 (1935).
6. Gregory, J. & Holmes, K. C. *J. molec. Biol.* **13**, 796–801 (1965).
7. Lauffer, M. A. *Trends biochem. Sci.* **9**, 369–371 (1984).
8. Finch, J. T. *et al. Nature* **212**, 349–350 (1966).
9. Vainshtain, B. K. *et al. J. molec. Biol.* **188**, 49–55 (1986).
10. Cohn, E. J. *Physiol. Rev.* **5**, 349–437 (1925).

tions with both the fungal proteinase and catalase is that they indicate simple methods for the attainment of the condition of supersaturation which may eventually lead to crystallization. Furthermore, the crystals obtained by these methods (macrogravity?) were of high quality, which appears contrary to some of the arguments proposed for the advantages of microgravity for the production of crystals of biological materials.

R. Leberman

EMBL,
Grenoble Outstation,
20 Avenue des Martyrs,
38042 Grenoble, France

Regeneration in sea lilies

SIR — Examination of living and fossil specimens suggests that sea lilies (stalked crinoids) have been capable of arm autotomy, the voluntary breaking off of a limb, and regeneration since the Palaeozoic, but direct evidence for this suggestion has never been obtained^{1,2}. On the basis of experiments performed on feather stars (unstaked crinoids), crinoids were thought to need at least one arm and an intact aboral nerve centre to regenerate other arms³. We report here our findings that a sea lily can regenerate not only all its arms simultaneously, but also even its entire crown.

Nine specimens of a sea lily (*Metacrinus rotundus*) collected at 150 m depth were maintained at $13 \pm 1^\circ\text{C}$ in darkened aquaria. Five of the sea lilies autotomized their crowns (namely the arms, radial plates and visceral mass) between 3 days and 5 months after the start of culturing; the crownless part comprised the basal plates, infrabasal plates and the entire stalk. For two other specimens, the crowns were removed by cutting at the junction of the radial and basal plates.

Six of the seven specimens (four of the five autotomized ones and both dissected ones) that had lost their crowns regenerated them on the remaining stalks. A small depression was left after the loss of the crown in the central area at the top

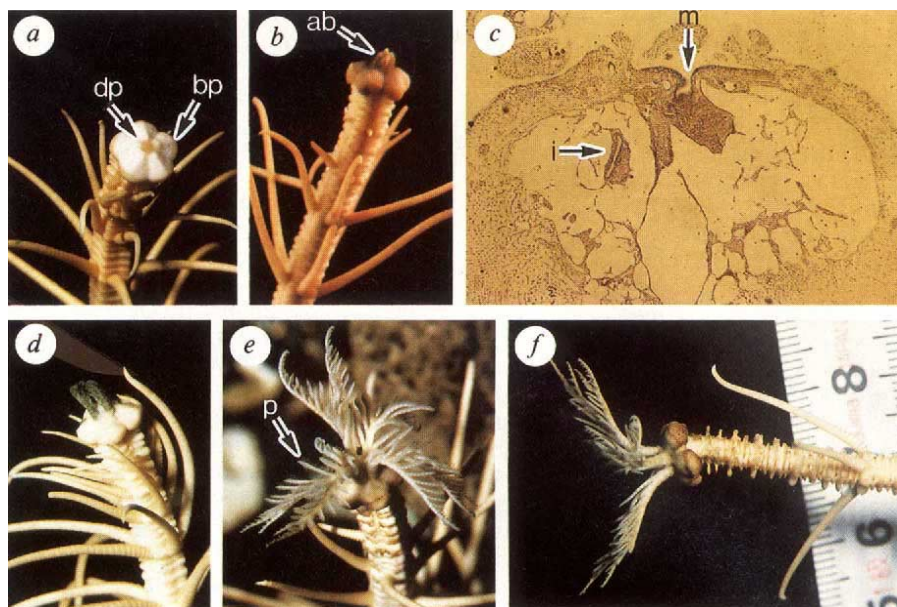


FIG. 1 Regeneration of the crown in *M. rotundus*. *a*, View from above the basal plates of a specimen 2 days after autotomy of the crown: dp, depression remaining after loss of the crown; bp, basal plate. *b*, Side view of the upper part of a specimen that had lost its entire crown by autotomy 99 days previously: ab, arm bud. *c*, Section approximately bisecting visceral mass of another specimen. The size of blastema of this specimen at the time of fixation was almost comparable to that of *b*: m, mouth; i, intestine. *d*, Oblique view of another specimen whose crown had been removed by dissection 141 days previously. *e*, Upper part of the stalk with regenerated crown in the same specimen as that in *b* 156 days later: p, pinnule which form podia. *f*, Side view of the same specimen as that in *b* 286 days later. The smallest graduation of the measure corresponds to 1 mm.

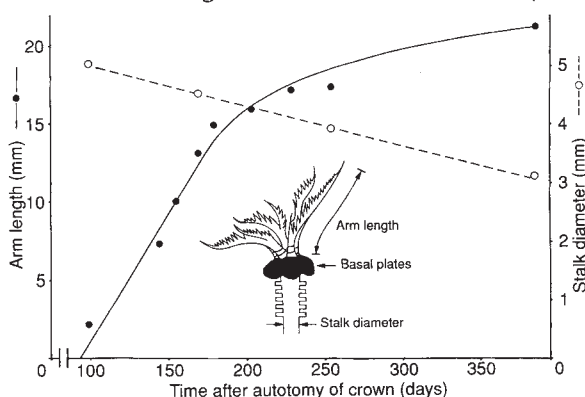


FIG. 2 Growth rate of an arm, and change in stalk diameter, during crown regeneration of a specimen corresponding to the one shown in Fig. 1*b*, *e* and *f*. Solid circles, arm length; open circles, stalk diameter.

plates (Fig. 1*a*). The depression was covered with a blastema within several days after loss of the crown. Tiny buds of five arms with several arm plates became distinguishable on the blastema within 3–7 months after crown loss (Fig. 1*b*). The visceral mass including mouth, intestine, anus and chambered organ has been completed by this stage (Fig. 1*c*). The view of the crown became conspicuous within 4–8 months (Fig. 1*d*). Many pinnules (branchlets of arm) differentiated along the oral side of each arm, and a fringe of podia was formed (Fig. 1*c*). The first arm branching (Fig. 1*d*, arrow) occurred following the shedding of the distal part of each arm. The growth rate of the longest arm in the regenerating crown was almost linear (0.17 mm per day) from the appearance of the arm buds to a length of 1.5 cm, then declined

and the arm reached 2.1 cm at 385 days post-autotomy (Figs 1*f*, 2). No food was supplied during the time the specimen was regenerating the crown.

The diameter of each columnal (stalk plate) in the uppermost 5 cm of the stalks decreased gradually while the specimens were regenerating their crowns (Fig. 1*f*). The diameter of the thinnest part of the stalk decreased from 5.0 mm at 99 days to 3.3 mm at 385 days post-autotomy at 1 cm below the basal plates (Fig. 2).

Our results indicate that the uppermost part of the sea-lily stalk, including the basal plates, has the capacity for regeneration of the entire crown. The aboral nerve centre, indispensable for arm regeneration in living feather stars³, is located at the oral end of the stalk and within the circle of basal ossicles in sea

- Emson, R. H. & Wilkie, I. C. *Oceanogr. mar. Biol. a. Rev.* **18**, 155–250 (1980).
- Meyer, D. L. & Ausich, W. I. in *Biotic Interactions in Recent and Fossil Benthic Communities* (eds Tevesz, M. J. S. & McCall, P. L.) 377–428 (Plenum, New York, 1983).
- Breimer, A. & Lane, N. G. in *Treatise on Invertebrate Paleontology Part T, Vol. 1* (eds Moore, R. C. & Teichert, C.) 316–347 (Geological Society, Boulder, Colorado, 1978).
- Grimmer, J. C., Holland, N. D. & Messing, C. G. *Mar. Biol.* **81**, 163–176 (1984).
- Brower, J. C., Lane, N. G. & Rasmussen, H. W. in *Treatise on Invertebrate Paleontology Part T, Vol. 1* (eds Moore, R. C. & Teichert, C.) 244–263 (Geological Society, Boulder, Colorado, 1978).

lilies⁴. This location supports the idea that the basal plates and a sufficiently long part of the stalk are crucial to regeneration of the entire crown in sea lilies. The function of the stalk in sea lilies is to elevate the crown above the substrate⁵. Our results indicate that the nutrients and materials from the stalk, especially from the upper 5 cm, are used for crown regeneration.

Shonan Amemiya

Misaki Marine Biological Station,
University of Tokyo,
Koaji, Miura-shi,
Kanagawa 238-02, Japan

Tatsuo Oji

Geological Institute,
University of Tokyo, Hongo 7-3-1,
Bunkyo-ku, Tokyo 113, Japan

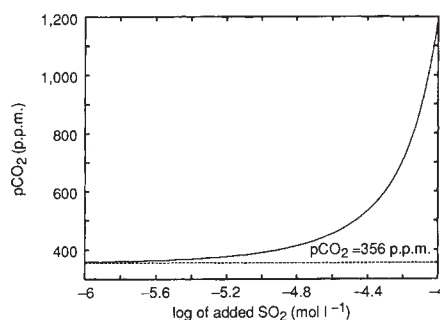
Disposal of SO₂ in sea water

SIR — Radojević and Tressider¹ confirm that gaseous SO₂ easily dissolves in sea water². They propose that seawater scrubbing may be used to reduce the anthropogenic emissions of SO₂, and that SO₂ could be injected "directly into an ocean current with no perceptible effect on seawater pH or composition". We find this statement misleading — in practical applications the injected SO₂ could lead to outgassing of oceanic CO₂, and thus contribute to further increase in the greenhouse effect.

The acidification of sea water as the injected SO₂ oxidizes to sulphate results in a shift in the inorganic carbon system; the carbonate and bicarbonate concentrations decrease whereas the concentration of dissolved CO₂ increases. The partial pressure of CO₂ as a function of only minor amounts of added sulphur dioxide is shown in the figure. The partial pressure of CO₂ (pCO₂) is 3 p.p.m. above its equilibrium value for an excess sulphate concentration of 10⁻⁶ mol l⁻¹, and a doubling of the equilibrium pCO₂ value occurs for 6.3 × 10⁻⁵ mol of added SO₂ l⁻¹.

If water with excess sulphate concentration greater than 10⁻⁶ mol l⁻¹ is in contact with the atmosphere, rapid outgassing of CO₂ would occur due to the increased pCO₂. To keep the concentration of the added sulphur below 10⁻⁶ mol l⁻¹, the 4 × 10³ g SO₂ s⁻¹ emitted from a 2-GW power plant¹ would have to be dissolved in a water volume of at least 6 × 10⁴ m³ every second. Most ocean currents are characterized by flow velocities well below 1 m s⁻¹, so this option requires an enormous diffuser device in order to dilute the injected SO₂.

Most ocean currents are characterized by flow velocities well below 1 m s⁻¹, so this option requires an enormous diffus-



Partial pressure of CO₂ in sea water as a function of added SO₂. The ambient sea water has a temperature of 10°C and salinity of 35 PSU, the total dissolved inorganic carbon content is 2.12 × 10⁻³ mol l⁻¹, and the alkalinity is 2.35 × 10⁻³ eq l⁻¹. This gives an equilibrium pCO₂ value of 356 p.p.m. The boric acid system has been included in the computations.

er device in order to dilute the injected SO₂.

For SO₂ disposal to be a practical option, currents leading from shallow water into the deep ocean are required. There are only a few natural sinking ocean currents, an example being the Mediterranean overflow into the Atlantic. However, negatively buoyant plumes may be created by dissolving CO₂ in sea water at elevated pressures^{3,4}, and thereby transport the water to the deep ocean. There might be a possibility for using a similar scheme for SO₂ or a combination of CO₂, SO₂ and other gases which dissolve in sea water, provided the gases have the property that they increase the density of the discharged water.

While disposal of flue gases into the ocean may have a potential to reduce atmospheric emissions, careful studies of the biological and environmental consequences of each gas are required before disposal can be implemented.

Helge Drange

Peter M. Haugan

Nansen Environmental and Remote
Sensing Centre,
Edvard Griegs vei 3A,
N-5037 Solheimsviken, Norway

SIR — Although the suggestion of Radojević and Tressider¹, that combustion emissions be piped below the sea to absorb SO_x and NO_x, may be attractive chemically, it does not hold water energetically. The engineers who design exhaust and cleaning systems for flue gases have to pay serious attention to the energy costs. Industrial electrostatic precipitators and bag filters currently in use for removing particles might involve a pressure drop of 1.47 kPa (150-mm water gauge).

Assume that the outfall of such a system has got to be far enough off shore

to give it a depth below the surface of, say, 10 m. A 2,000-MW electricity generating station would have a total gas emission rate of 2,000 m³ s⁻¹. Hence the power used to overcome the pressure drop due to the depth of water would be dV/dt × dP, or 2,000 × (10 × 1,000 × 9.81), around 200 MW. The station would thus use 10% of its own power output to clean its own flue gases. Add to this waste the probability that less than 100% of the gases would be absorbed, the thermal input from flue gases at 120°C and the creation of foaming toxic regions of sea with consequent damage to marine flora and fauna — surely this could never be the best environmental option.

J. J. Colls

Department of Physiology and
Environmental Science,
University of Nottingham,
School of Agriculture,
Sutton Bonington, Loughborough,
Leicestershire LE12 5RD, UK

RADOJEVIĆ REPLIES — Drange and Haugan, and Colls, raise some important issues following from our report¹. Whether our proposal is economically viable depends on the value one attaches to the environment and its protection as well as any energy "penalty". Currently popular limestone-based desulphurization techniques involve enormous expense, for example.

We did not discuss desorption of CO₂ from acidified solutions, although this is a possibility. We have carried out some preliminary experiments in which CO₂ (13% in air) was bubbled into sea water. The sea water has some CO₂-absorbing capacity although this was lower than that for SO₂. Absorption of CO₂ is a reversible process, unlike absorption of SO₂ which is rapidly oxidized to sulphate, a stable product. In real applications of the proposed desulphurization method the extent of CO₂ desorption will depend on dilution factors, depth of discharge and other parameters, and will be limited to the locality of discharge. Once the effluent is diluted to background levels desorption of CO₂ will become unimportant.

Methods involving direct discharge of flue gases into sea water are still in the development stage and until actual field trials are carried out some of the questions raised above will remain unanswered. At first it may be more practical to apply the technique to smaller industrial plants until some of the technical problems are resolved. The case of the 2,000-MW plant in our report was used only as an illustration and we are still some way off from applying the method to such large plants.

Sea water is a far more suitable recipient for combustion-generated pollution

tion than the atmosphere. CO₂, SO₂ and HCl are readily converted into soluble ions which are natural constituents of sea water. In fact the composition of sea water may reflect absorption of these gases by the oceans from the primordial atmosphere and volcanic eruptions over the geological timescale.

The work carried out to date demonstrates that sea water is a suitable recipient for SO₂ pollution. Clearly more research is needed to find suitable solutions to the problem of CO₂ emissions. It should, however, be borne in mind that the two pollutants are usually present together and a solution to both problems using one method would be desirable.

Miroslav Radojević

Department of Chemistry,
University of Brunei Darussalam
Bander Seri Begawan,
Negara Brun Darussalam

1. Radojević, M. & Tressider, D. A. *Nature* **356**, 391 (1992).
2. Liess, P. S. & Slater, P. G. *Nature* **247**, 181–184 (1974).
3. Baes, C. F. Jr, Beall, S. E., Lee, D. W. & Marland, G. in *Interactions of energy and climate* (eds Bach, W., Pankrath, J. & Williams, J.) 495–519 (Reidel, Dordrecht, 1980).
4. Haugan, P. M. & Drange, H. *Nature* **357**, 318–320 (1992).

Chromosomal instability

SIR — Kadhim *et al.*¹ studied clonal descendants from murine haematopoietic stem cells irradiated with α -particles from plutonium-238 (linear energy transfer, LET = 121 keV μm^{-1}). They observed a high frequency of nonclonal chromosomal aberrations after 10–13 cell divisions following the initiation of proliferation, even though this proliferation was clonal. Most of the anomalies were of the chromatid type, resulted in breakages and were assumed to affect chromosomes or chromosome sites at random. The authors interpreted their results as a *de novo* occurrence of transmissible chromosomal instability, and assumed that this instability, which was not observed after X-ray irradiation under the same conditions, characterized the biological effects of high LET radiation.

We began our investigation of radio-induced chromosomal anomalies by characterizing the complex rearrangements induced by heavy ions in human lymphocytes², and then focused on the transmissibility of radioinduced chromosomal rearrangements in human dermal fibroblasts (see table). For this purpose, fibroblasts were irradiated by neon ($E=10.74$ MeV per a.m.u., LET=386 keV μm^{-1} ; E is energy), argon ($E=10.52$ MeV per a.m.u., LET=1,207 keV μm^{-1}) and lead ($E=9.5$ MeV per a.m.u.,

CYTOGENETIC ABERRATIONS IN HUMAN FIBROBLASTS EXPOSED TO HEAVY IONS

Radiation	Fluence (particle per cm ²)	Dose (Gy)	Subculture	Metaphases with aberrations	Metaphases with at least 1 dicentric	Metaphases with clonal rearrangement	Metaphases with clonal rearrangement and at least 1 dicentric
Neon	10 ⁶	0.62	P20	22/23	16	0	0
	2×10 ⁶	1.24	P25	17/19	12	0	0
	4×10 ⁶	2.48	P25	12/12	1	3	4
Argon	10 ⁶	1.93	P25	11/12	7	0	0
	2×10 ⁶	3.86	P25	21/22	8	10	0
	4×10 ⁶	7.72	P23	20/20	3	14	2
Lead	2×10 ⁶	44	P20	15/24	9	0	0

Human dermal fibroblasts from a normal donor were irradiated at confluence at the G.S.I. (Darmstadt). Fluences ranged from 10⁶ to 4.10⁶ particles per cm², which corresponded to the passage of 0.4 to 1.6 particles through the cell nucleus (equatorial area, 40 μm^2). They were then cultured up to the twentieth or twenty-fifth subculture (about 40 cell divisions). R-banded karyotypes were established for each of the five subcultures. Almost 25% of the chromosomal breakpoints detected at the fifteenth to twenty-fifth subcultures were located on chromosome 13. See text for radiation exposures.

LET=13,600 keV μm^{-1}). At the earliest passages studied (1–3), a high percentage of metaphase cells carried chromosomal aberrations directly induced by irradiation, most of which disappeared after a few subcultures. At passages 5–10, almost all karyotypes were normal except in the cultures irradiated by neon ions, in which transmissible chromosomal rearrangements persisted. This was obviously due to the low dose of irradiation delivered by these ions (0.62–2.48 Gy).

Around the fifteenth subculture, a high percentage of dicentrics, resulting from end-to-end translocations, was observed. The location of the corresponding breakpoints were not random but specific, as they were found mainly in the two telomeric regions of chromosome 13, and were also observed, in decreasing order of frequency, in that of chromosome 16, of the short arm of chromosome 1 and of other chromosomes.

Therefore, this nonrandom chromosomal instability, which affected more than 50% of metaphase cells, can be considered as a late consequence of radiation by high LET particles. It further led to clonal aberrations which were frequently unbalanced, and in particular to the loss of chromosome 13. Consequently, a period of chromosomal instability induced by heavy ions, followed by the formation of clones with unbalanced karyotypes, could constitute early steps towards cellular transformation.

These results for normal human fibroblasts confirm those of Kadhim *et al.* for murine haematopoietic stem cells with respect to the induction of chromosomal instability after high LET irradiation. However, we studied the effects of particles with a wide range of high LET radiation (386–13,600 keV μm^{-1}), where-

as Kadhim *et al.* used only Pu-238 (LET=121 keV μm^{-1}). In addition we found, in contrast to Kadhim *et al.*, that the chromosomal instability, mainly reflected by the appearance of dicentrics, which are nontransmissible chromosomal rearrangements, does not appear at random but affects certain specific telomeric regions. We further observed that this instability persisted from at least the fifteenth to the twenty-fifth passage; and that it led to clonal chromosome imbalances and rearrangements, as frequently observed in human solid tumours.

These results may help to explain why high LET radiations are so potent in causing cell transformation³ and radiocarcinogenesis⁴.

Laure Sabatier

Bernard Dutrillaux

Commissariat à l'Energie Atomique,
PO Box 6,
92265 Fontenay-aux-Roses,
France

Maria Berta Martin

Laboratório Nacional de Engenharia e
Tecnologia Industrial,
DPSR/SPR, Estrada Nacional – 10,
2675 Sacavém,
Portugal

1. Kadhim, M. A. *et al.* *Nature* **335**, 738–740 (1992).
2. Sabatier, L., Al Achkar, W., Hoffschir, F., Luccioni, & Dutrillaux, B. *Mutat. Res.* **178**, 91–97 (1987).
3. Yeng, T. C., Craise, L. M., Mei, M. T. & Tobias, C. A. *Radiat. Res.* **104**, S177–187 (1985).
4. Fry, R. J. M., Powers-Risius, P., Alpen, E. L. & Ainsworth, E. J. *Radiat. Res.* **104**, S188–195 (1985).

Scientific Correspondence

Scientific Correspondence is intended to provide a forum in which readers may raise points of a scientific character. They need not arise out of anything published in *Nature*. In any case, priority will be given to letters of fewer than 500 words and five references.

Computing beliefs

Igor Aleksander

Artificial Believers: The Ascription of Belief. By Afzal Ballim and Yorick Wilks. Lawrence Erlbaum: 1991. Pp. 285. £25, \$42.50.

IN about the fortieth year of its existence, artificial intelligence (AI) may just be beginning to show signs of maturation. Gone are the heady days when AI workers tried to outdo one another with public statements about the potential intelligence of computers. Many will remember Ed Fredkin, the manager of the AI laboratory at the Massachusetts Institute of Technology, turning to a television camera in 1983 and saying: "There is no principle of science or engineering that prevents us from making computers that are infinitely smarter than ourselves They will talk to us only to amuse themselves and so, in some sense, keep us as pets."

By comparison, the AI described by Afzal Ballim and Yorick Wilks has a ring of truth, through its detachment from the pretence that it makes machines 'smart'. They define AI as "the writing and running of programs to do things that are done in AI laboratories". In this case, the thing done in AI laboratories is the writing of programs that, in some way, could be said to represent 'beliefs' (as opposed to older pursuits that might have represented, say, the expertise needed to play chess or to solve simple problems). This definition, although broadcasting some obvious cynicism, may be appropriate and correct. It says that AI is an abstract study which, like the proving of theorems in pure mathematics or deciphering ancient liturgical texts, is a perfectly legitimate intellectual academic pursuit.

Ballim and Wilks genuinely want to engage the reader in work that is intrinsically interesting. The centrepiece of the book is a computer program (ViewGen, for viewpoint generator), which the authors describe in some detail. There are probably no more than a few dozen readers in the world who will fall avidly on this detail and feel that they have learned something new. The authors are fully aware of this, saying that "the hardest liners, who want nothing but system details and are confused by prose, can begin at chapter 3". Chapter 3, indeed, is a description of the ViewGen program, whereas chapters 1 and 2 set the scene in terms of the adopted methodology and its relationship to philosophy and psychology.

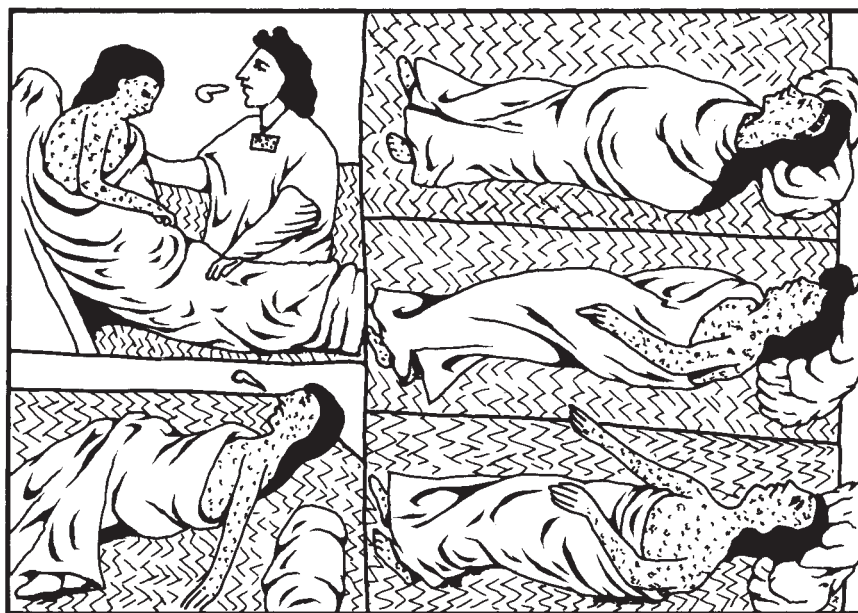
In these introductory chapters, the new, isolationist view of AI is presented with firm statements that the book is not

a contribution to philosophical discourse about the nature of 'beliefs'. Anything as complex as a belief (and beliefs about other people's beliefs can get awfully complex) should be described by a logical structure that is clear and unmistakable. At a trivial extreme, one could take the view that when the computer prepares one's telephone bill for £92.75, it actually believes that this is what one owes. At the other extreme, it is possible to have a discussion (likely to be inconclusive) about whether a computer could ever believe in God. To make the AI discourse meaningful, Ballim and Wilks avoid these extremes and base their work on the nature of linguistic interactions that, somehow or other, assume a belief. To be clearer, consider a statement that I might make to a student at the end of the final examination: "I'm

sure that you will be pleased to hear that both you and your friend Jane got first-class honours degrees". This linguistic interaction contains assumptions about the listener's beliefs, such as that getting a degree is a good thing and that Jane thinks so too.

How would one write a program that could contain such beliefs? What does it mean for a program to contain any beliefs at all? Answering such questions is what this corner of AI is about. The authors take a path that is centred on 'ascription', that is, the ability to ascribe beliefs to others without direct evidence and through the use of 'default reasoning'. The latter relates to general facts about the world that the system would have to know and understand. Take, for example, the following two sentences: "When Tommy wakes up on Christmas Day he'll rush to the fireplace to look for his stocking. He'll be pleased to find that nobody has given him a book of fairy tales, which he dislikes intensely". Here the speaker makes assumptions about Tommy's knowing general facts about Christmas and Christmas stockings — this is default reasoning. Any theory of beliefs must account for this. But it must also somehow be capable of including

Old World disease



When Columbus arrived in America in 1492 there were already about 100 million native inhabitants. Within decades most of these people were dead, victims of Christian fanaticism and sheer barbarity, but also of diseases imported by colonists. Smallpox, measles, influenza, bubonic plague, yellow fever, cholera and malaria — all unknown in the Western Hemisphere before 1492 — killed at least half the population of the Aztec, Maya and Inca civilizations shortly before their overthrow. Shown here are smallpox victims as portrayed in the Florentine Codex, a history of the conquest written by Aztec scribes for Friar Bernardino de Sahagun in the 1550s. A powerful account of the European conquest of native America is given by Ronald Wright in *Stolen Continents: The Indian Story*. Published by John Murray, £19.95.

very specific beliefs, such as Tommy's not liking fairy tales. The ViewGen program is all about logical methods for holding and transferring default and specific beliefs.

The authors are aware that sciences of artificial reasoning are going through interesting times as a result of the revival of 'connectionism' — representations of knowledge and reasoning in artificial models of the neural structure of the brain. The difference between connectionism and older logical AI models lies in the fact that connectionists try not only to include models of reasoning but also to provide explanations of how much prowess could be learned. Ballim and Wilks claim to "tread an uneasy path" on this issue by not commenting on the acquisition of beliefs at all. Their science is limited to a belief structure generated by a programmer and not the organism itself. Despite being convinced of the power of neural modelling, I do not find this reticence unreasonable. It merely leaves open the question of how a distributed, adaptive mechanism such as the living brain could get itself into a state where it can perform the logical operations so clearly set out in Ballim and Wilks' theoretical model.

Having been impressed by the sober line of logical argument taken through most of the book, I was surprised by the last chapter, entitled "Practical and Social Matters". Here the authors take the enormous leap of suggesting that machines might develop viewpoints and beliefs of their own (having explained earlier that this possibility was not implied in their system). This is put at the door of the natural tendency for humans to humanize mechanical behaviour. For example, I might say: "My car's being stubborn — it doesn't want to start". People might attribute the power of human judgement to machines with simulated beliefs, and this could clearly be a bad thing. Ballim and Wilks conclude, somewhat astonishingly, that laws will have to be passed that constrain the activities of owners of machines that contain belief programs, in much the same way as there are laws that constrain the ownership of dogs. It is the owner's responsibility that the beliefs held in these machines should not harm anyone. Despite this seemingly unnecessary excursion into speculation, the book should be read by anyone who has wondered what a 'belief' might be. For the nonspecialist, the technical detail can easily be ignored without diminishing the undoubtedly stimulating contents of this book. □

Igor Aleksander is in the Department of Electrical and Electronic Engineering, Imperial College of Science and Technology, Exhibition Road, London SW7 2BT, UK.

Getting an ought from an is?

Stuart Sutherland

Science as Salvation: A Modern Myth and its Meaning. By Mary Midgley. Routledge: 1992. Pp. 239. £25, \$25.

SCIENCE is having a bad press. Orthodox medicine is shunned in favour of homeopathy or aromatherapy; the environmentalists blame pollution on science rather than on its misuse; nobody likes the atomic bomb; and now we have a philosopher, Mary Midgley, and a journalist, Brian Appleyard, telling us in remarkably similar books that science has stolen man's soul and left him without religion, aims or values. (For reviews of Appleyard's book, *Understanding the Present*, see *Nature* 356, 729 and 357, 29; 1992.)

Most religions serve two functions. They satisfy man's desire for explanations and they provide a code of conduct. Science has taken over the first role if only because it is better at it: nowadays few can regard the Garden of Eden as anything other than an entertaining myth. Because they confused religion's double role, clerics were upset by the Copernican theory and even more upset by Darwinism, while the Church of England is at the moment busy arguing about the truth of the Resurrection, instead of getting on with the more important task of saving souls. Midgley fails to distinguish religion's two functions. She appears to resent science's intrusion on religious myths, but her main argument is that science is illegitimately taking over the provision of moral and spiritual guidance. She has assiduously collected some foolish quotations from scientists, which, so she claims, demonstrate that they are attempting to subvert people by imposing inappropriate values based on scientific discoveries. Her method is not very convincing: of course scientists sometimes say silly things; so for that matter do philosophers. She recognizes, only to ignore the fact, that "There are indeed irritating people, both in the arts and in the sciences, who often argue for positions that they do not take seriously." She continues in her rather schoolmarmish way, "They are generally recognised as a pest." So much for Oscar Wilde.

Nor does she distinguish speculation from prophecy. It is hard to believe that John Barrow and Frank Tipler did not have their tongues in their cheeks when they forecast that the eventual collapse of the Universe would be prevented by beings shovelling matter down black holes and that the Universe would reach

Omega Point, at which time life will have spread everywhere and we will know "everything there is to be known". Physicists should attach a health warning to such statements to stop serious-minded philosophers ascribing too much weight to them. Indeed, the other day I overheard a physicist threatening to tip Dr Midgley down a black hole, along with Appleyard, but it turned out he was only joking.

From time to time Midgley gets something right. For example, she rejects the 'strong anthropic' principle, which is based on a teleological argument. People could not have evolved if the constants of physics were anything other than they are. Therefore, these constants exist solely for the purpose of producing us. One might just as well argue that if the presence of a log in a fast-flowing river saves someone from drowning, then that was the purpose of the log. Again, she points to an obvious problem for Freeman Dyson's prediction that people will colonize the Solar System, namely, the difficulty of overcoming the massive gravitational force of most planets. She should, however, bear in mind that 50 years ago few would have predicted man's ascent to the Moon, let alone the invention of the digital computer.

Several scientists, such as J. D. Bernal and J. B. S. Haldane, became crackpots when they left their laboratories. Midgley uses their armchair utterances to show that science is making a takeover bid for human values. The only discernible values that have anything to do with science are the search for understanding (which Midgley wilfully misinterprets as the search for disconnected pieces of information) and the attempt to make intelligent beings of some sort survive into the distant future or possibly for all time (whatever that means). These aims do not appear to be particularly discreditable and it is surely obvious to all that although science might possibly provide the means to their achievement, it cannot of itself justify them.

Oddly, neither Midgley nor Appleyard discusses what many would take to be science's biggest threat to human values — determinism, with which Tennyson and many other Victorians wrestled. If the world on a human scale is largely determined, what room is left for free will, let alone the formation of values? Neither quantum theory nor chaos lends succour, for having our activities governed probabilistically is no more satisfactory than having them fully determined.

The worst feature of Midgley's book is that, like Appleyard, she is prone to hysterical exaggeration. Do people have

fewer or worse values now than in Ancient Rome or in Mediaeval times? Even in Queen Victoria's golden days, long before modern cosmology began, policemen were unable to patrol the streets of London without carrying firearms and being in pairs. At a time when only one-third of British 13-year-olds believe the Earth goes round the Sun, how many people have really had their values stripped away by science? Yet Midgley writes: "The world's new position is often felt to have astonishing moral and metaphysical consequences, such as that truth or justice has become a meaningless concept — sometimes, even more strangely, the psychological consequence that all human kindness is an illusion." No evidence is offered for this extraordinary statement. Since most people remain honest, kind and decent, one wonders what company she has been keeping: surely her philosopher colleagues cannot be that vicious.

Leaving aside the speculations of philosophers and cosmologists, it is unclear why values should be in the least affected by discoveries in physics. One cannot, as the saying goes, "get an ought from an is". Values and morality concern how we chose to spend our time, how we treat others and, one hopes, how far we try to avoid damage to future generations through maltreating the Earth. What occurs in other galaxies is irrelevant — unless, of course, intelligent beings were found there, in which case we might have to think hard about how to treat them. Midgley does not even stop to ask whether if there really is a malaise in the Western world, it might be caused by factors other than science, such as the environment of inner cities, the breaking down of the traditional roles of the different classes, increased tolerance for divergent behaviour or even television. Does she really suppose that those yobboes on Tyneside were prompted to engage in car burn-ups as a result of reading Stephen Hawking or that the violence of the Ku Klux Klan was inspired by too close an acquaintance with relativity theory? □

Stuart Sutherland is in the Laboratory of Experimental Psychology, University of Sussex, Brighton BN1 9QG, UK.

■ The relationships between modern science and society are explored in two other recently published books aimed at a popular audience. In *The Creative Moment*, Joseph Schwartz, coauthor of *Einstein for Beginners*, looks at "how science made itself alien to modern culture" (HarperCollins/Jonathan Cape; \$25, £16.99). In *The Science Gap*, Milton A. Rothman sets about "dispelling the myths and understanding the reality of science" (Prometheus, \$24.95).

Under the net

Christopher W. Petersen

The Ecology of Fishes on Coral Reefs. Edited by Peter F. Sale. *Academic*: 1991. Pp. 754. \$75.00, £46.50.

CORAL reefs are one of the most diverse and productive ecosystems in the world (and also one of the most threatened). A main component of reef fauna are fishes, and those inhabiting reefs form one of the most assorted and abundant assemblages of vertebrates to be found anywhere in the world. Yet ten years ago, reviews of the ecology of fishes on coral reefs were mainly limited to arguments



I. F. Took/Biotofoto

Juvenile grunts and blackbar soldierfish on a 3-year-old reef in the Caribbean.

over the relative roles of competition and chance in determining the sizes of adult fish populations. That time is over. This outstanding book marks the maturation of the field as a dynamic area of study that has successfully shed its simple early hypotheses for a rich blend of experimentally based research programmes. The volume will quickly become required reading and the main review source for reef-fish biologists, and also an essential work of reference for other interested ecologists.

Peter Sale has succeeded in bringing together a wide range of different viewpoints under one cover. The book's 20 chapters are written by a list of authors that reads like a *Who's Who* of reef-fish ecologists, with a blend of US workers planted firmly in the Caribbean, and Australian and New Zealand workers focusing their efforts in the Indo-Pacific, particularly on the Great Barrier Reef.

By studying different species in different places, the authors often see different factors regulating population size and community structure; so it is hardly surprising that they sometimes disagree

on the importance of various processes in the ecology and evolution of reef fishes. But workers now acknowledge these differences and attempt to put their findings into perspective.

Sale has done an excellent job of balancing the contributions. The book is divided into six sections covering basic material such as the visual world, morphology, evolution and palaeontology of reef fishes; trophic relationships; larval and juvenile ecology; reproductive and life-history patterns; community organization; and fisheries and their management. Palaeontology and the management of fisheries are subjects that are not generally discussed in journals on reef-fish ecology, so the book will certainly broaden the horizons of a lot of workers in the field.

Many of the authors try to frame their findings in the wider picture of the patterns and processes they see in reef-fish ecology. Several look at factors that limit local population size, but many focus on coevolution in reef communities, from patterns of herbivory and plant defences to biogeographical and palaeontological patterns. Embedded in the chapters are plenty of research programmes that will inspire those just beginning to study reef fishes.

An area not covered that deserves more attention is the genetic structure of populations of reef fishes. Despite the meagre amount of data so far published on the subject, estimates of genetic differentiation among populations should eventually provide useful insights into biogeographical trends, fisheries management and the consequences of inter-specific differences in larval biology.

Sale first shook up the ideas of many biologists in the early 1970s when he proposed that niche partitioning was not important for reef fishes. This non-MacArthurian view of community structure, with stochastic processes dominating deterministic ones (Sale's lottery hypothesis), started a debate that continues in this book. New ideas about factors that limit population size, from the arrival of very few planktonic larvae at a reef (pre-recruitment limitation) to reef-based factors such as post-recruitment predation, challenge the assumptions of adult space limitation made by Sale and his predecessors. Just as Sale should be commended for his early questioning of ecological dogma, he should be congratulated on pulling together an excellent set of contributions. Compiling books on the ecology of a particular group of organisms is itself a bit of a lottery, but with this one Sale has hit the jackpot. □

Christopher W. Petersen is at the College of the Atlantic, 105 Eden Street, Bar Harbor, Maine 04609, USA.

The anatomy of the Sun

J. B. Zirker

Solar Interior and Atmosphere. Edited by A. N. Cox, W. C. Livingston and M. S. Matthews. *University of Arizona Press: 1991. Pp. 1,416. \$65, £57.95. Distributed by Eurospan in Europe.*

OUR Sun proves the rule that the closer you look at anything in nature, the more complicated it becomes. For at least 40 years, astronomers have sought satisfactory explanations for some of the most obvious solar phenomena. Why does the Sun's equator rotate faster than its poles? Why is its outer atmosphere so incredibly hot? What causes its 11-year cycle of activity? What, exactly, is a sunspot? How, indeed, does the Sun generate the huge explosions that rattle the Earth's magnetic field?

During the past decade, dramatic progress has been made in answering these questions. We now have better equipment on the ground and in space, and a deeper knowledge of plasmas and elementary particles, coming in part from the controlled fusion programme. Solar problems now attract some of the best minds in astrophysics, and there is a growing realization that we have much to learn from other stars similar to the Sun. Solar astronomy has grown to become solar physics, and new generations of physicists have answered the call.

All these developments have been exciting, but devilishly difficult to keep up with. At last, this blockbuster of a book has appeared to summarize the subject and point the way to the future. It is one of two volumes on the Sun in the University of Arizona Press Space Science Series*. Two years in the making, with an international cast of 101 authors and around 1,400 pages, this monumental volume consists of a well-integrated collection of topical reviews of many of the principal aspects of solar phenomena. Readers need a graduate's competence in physics and mathematics and a good grounding in solar observations. For the most part, the book is written clearly and scholarly, but not pedantically. Each chapter has an abstract and a summary. There is a useful glossary and a comprehensive up-to-date list of references. A fine sense of the excitement of research comes through in several chapters, and, understandably, there is a flavour of controversy, even contradiction.

* The companion volume is *The Sun in Time* edited by C. P. Sonett, M. S. Giampapa and M. S. Matthews, an interdisciplinary review of the Sun's evolution written by 83 collaborating authors. \$60.

The book is divided into five main sections covering the physics of the solar interior, including convection theory, structural modelling and dynamo theory; solar-oscillation observations and theory; the structure of the solar surface; solar magnetic activity in all its guises; and the Sun as a star.

The chapters on helioseismology are particularly enjoyable. This field is only about 15 years old and is developing into a marvellous tool for exploring the deep interior of the Sun. The subject is elegant, both in its experimental and theoretical aspects. More importantly, the study of oscillations is guiding attempts to understand how the Sun generates magnetic fields and why the solar cycle behaves as it does.

One cannot fail to be struck by the shift of emphasis that has occurred since G. Kuiper's compendium *The Sun* was published in 1953. At that time, radiative transfer, spectral-line analysis and nonthermodynamic equilibrium studies were at the cutting edge of research. Now the focus is on hydrodynamics,

hydromagnetics and the interaction of large- and small-scale systems. Indeed, the basic mechanisms underlying flares, the heating of the corona and the solar wind, all involve solar magnetism at many spatial and temporal scales.

The authors present current observations without much discussion of the instruments that made them possible. This is something of a shortcoming, because there has been a revolution in solid-state detectors, X-ray telescopes, imaging spectroscopy and computer analysis of huge data sets. But even in a tome of this size, something has to be omitted.

For all our increased knowledge about the Sun's behaviour, new questions keep popping up. Where are all the missing neutrinos? How is the solar wind driven? There is, fortunately, no end to the fascinations of this seemingly prosaic star. □

J. B. Zirker is at the National Solar Observatory, Sunspot, New Mexico 88349, USA.

Difficult superconductivity

A. C. Rose-Innes

High-Temperature Superconductivity: An Introduction. By Gerald Burns. *Academic: 1992. Pp. 199. \$19.95, £13.*

GERALD Burns admits to having found this book difficult to write. I am not surprised; it has been estimated that 20,000 papers were published on high-temperature superconductivity in the five years after its discovery. To read their titles alone would probably take about 30 hours.

Burns is surely over-optimistic when he says that there has been a "remarkable rate of progress" in the study of high-temperature superconductors; my impression after reading this book is that there is in fact still little understanding of these materials and few areas in which they have been successfully applied. This situation will probably persist until reasonably large single crystals are readily available for experimentation. I am reminded of the similar confusion over the properties of semiconductors in the early 1950s before the advent of single crystals of germanium and silicon.

It is not clear to me for whom this book is written. It is a curious mixture of the very elementary, such as a basic introduction to perfect diamagnetism, and of advanced concepts, such as a detailed discussion of Ginzburg-Landau theory. But it is not really suitable for novices such as undergraduates. Indeed,

does Burns really believe that the topic of high-temperature superconductivity involving highly anisotropic extreme type II superconductors, the understanding of which is defeating established experts in the field, is suitable for "students taking their first solid-state physics course"?

For experts, by contrast, the information on high-temperature superconductors is diluted by presentations of elementary superconductivity, even though the author states that readers are assumed to have this background knowledge already.

Despite the inclusion of elementary and sometimes irrelevant material, the book provides a much needed review of the field up to 1991. It is to a large extent a catalogue of often conflicting experimental and theoretical results. In this respect, the summary sections will prove to be particularly valuable. No one can fully cover the vast amount of literature on high-temperature superconductivity — I know workers who feel so overwhelmed that they have given up attempting to read the literature at all. A review volume such as this will certainly be useful to them. Indeed, every researcher in high-temperature superconductivity would benefit from obtaining a copy. □

A. C. Rose-Innes is emeritus professor of physics and electrical engineering in the University of Manchester Institute of Science and Technology, PO Box 88, Manchester M60 1QD, UK.

Formation of an Archaean continent

Maarten J. de Wit, Chris Roering, Rodger J. Hart, Richard A. Armstrong, Cornel E. J. de Ronde, Rod W. E. Green, Marian Tredoux, Ellie Peberdy & Roger A. Hart

About 30% of the Earth is covered by continents, but only about 10 small kernels of these continents—known as Archaean cratons—are continental fragments formed before 2.5 Gyr ago. The Kaapvaal craton of South Africa, which formed and stabilized between 3.7 and 2.7 Gyr ago, is one of the oldest reasonably sized examples of these continental fragments. It consists of a mosaic of subdomains that have been welded together by processes similar to those of modern-day plate tectonics. The earliest subdomains may have owed their origin to the onset of efficient recycling from the Earth's hydrosphere into the mantle.

SOME of the most fundamental problems concerning the Earth's evolution relate to the first emergence of the continental lithosphere and to the associated tectonic and thermal processes then operating. There is as yet no consensus as to when and how the earliest continental regions were formed, and how much and at what rates continental lithosphere was extracted from the mantle in the Archaean¹. Although no continental rocks older than about 4.0 Gyr have been found², evidence from minerals (zircons) indicates that continents in some form may have existed as far back as 4.2 Gyr (ref. 3). The lack of preservation of continental regions older than 4.0 Gyr may indicate that earlier continents were unstable and were recycled back into the mantle, perhaps in response to intense extraterrestrial bombardment⁴. It is also possible that before 4.0 Gyr, continents on Earth were very rare. If so, we need to explain the cause of the sudden increase in continent formation after 4.0 Gyr, and how the Earth evolved from a state in which it was completely covered by oceans to its present configuration.

It is generally accepted that continents at present grow through a combination of subduction-related additions of new materials from the mantle and amalgamation of pre-existing lithospheric fragments of various sizes, often referred to as terrane accretion⁵. The growth of continents, however, is not synonymous with the concept of continental lithosphere growth: the latter refers to the net addition of new mantle-derived material to the continents, and thus takes into consideration accretion as well as recycling of continental material back into the mantle. Continental lithosphere growth remains a controversial topic, in part because there is disagreement about the rate of continental recycling, which has almost certainly varied greatly since the Archaean¹. Meaningful analysis of these processes is essential if we are to understand the formation of the earliest continental fragments.

The crust of the earliest continental fragments is predominantly a variable mixture of plutonic rocks with calc-alkaline-like chemistry and supracrustal rocks such as sediments and volcanic rocks, which are often green in appearance. The former are traditionally referred to as 'granites', and the latter as 'greenstones'; together, these rock sequences make up the 'granite-greenstone terrains' of the Archaean cratons.

Many Archaean cratons consist predominantly of granite-greenstone sequences formed in the late Archaean (2.5–3.0 Gyr ago). Those cratonic areas that have an older history have generally been extensively deformed and metamorphosed, and thus lost much of their original character. The Kaapvaal craton in southern Africa and the Pilbara craton of northwestern Australia are the only regions that have retained a substantial portion of relatively pristine mid-Archaean rocks (3.0–4.0 Gyr ago). Because of the enormous accessible mineral wealth of the Kaap-

vaal craton, this region (Fig. 1) has probably been more extensively investigated than any other Archaean craton.

The Archaean history of the Kaapvaal craton (or simply the craton) spans about 1 Gyr and can be conveniently subdivided into two periods, each of about the same length as the Phanerozoic period (see Table 1). The first period, ~3.7–3.1 Gyr ago, records the initial separation of the continental lithosphere of the craton from the mantle. This period of continental growth resembles plate tectonic processes occurring in modern-day ocean basins, but the difference is that in the mid-Archaean, these oceanic processes seem to have occurred in shallower water depths. The oldest oceanic rock sequences that have retained evidence for shallow submarine processes, like those

TABLE 1 First 1 Gyr of formation of the Kaapvaal continent

Kaapvaal shield formation*	500 million years of intraoceanic tectonics	Time (Gyr BP)
	First regional emergence of Kaapvaal continental lithosphere through intraoceanic obduction processes	3.7–3.2
	Western Pacific-style intraoceanic amalgamation of oceanic terranes	3.3–3.2
	Widespread within-shield melting, granite formation and chemical differentiation in the upper lithosphere	3.2–3.1
	<i>Kaapvaal shield stabilized by</i>	3.1
Kaapvaal craton formation*	500 million years of inter- and intra-continental tectonics	
	Regional extension of the Kaapvaal shield; passive continental margin and rift basins	3.1–2.9
	Cordillera-type accretion of composite terranes along north and west margin of Kaapvaal shield; intermontane and foreland basins	2.7–2.9
	Alpine or himalayan-type continent-continent collision between the Kaapvaal and Zimbabwean shields	2.68 ± 0.05
	Foreland extensional and impact-generated rifts	2.7–2.6
	<i>Kaapvaal craton stabilized by</i>	2.6

* Geologists refer to that portion of a continent which has attained stability and which has been little deformed for a prolonged period as a craton. The inner portion of a craton is referred to as a shield. Shields themselves are initially unstable and subjected to extensive deformation and internal chemical differentiation before cratonic rigidity is attained.

occurring at mid-ocean ridges, are found in the Barberton and adjacent granite-greenstone terrains of the southeastern regions of the craton, and within the Pietersburg greenstone terrain along the northern margin of the craton (see Fig. 1).

The second period of the craton's Archaean history (3.1–2.6 Gyr) records intra-continental and continental-margin processes: continental growth during this period occurred mainly through a combination of tectonic accretion of crustal fragments and subduction-related igneous processes in much the same way as has been documented along the margins of the Pacific and Tethys oceans since the Mesozoic. Examples are best observed in the northern parts of the craton, where the 'continental-shelf' terrain of the Central Limpopo mobile belt (Fig. 1a, inset A) has been tectonically juxtaposed between the granite-greenstone terrains of the Kaapvaal and Zimbabwean cratons (Fig. 1); and in the western margin of the Kaapvaal craton, which consists largely of late Archaean amalgamated granite-greenstone fragments. Evidence for tectonic processes similar both in style and time to those along the northern and western craton has also been found in the older central and southern parts of the craton, where these tectonic processes are closely related to formation of major late Archaean sedimentary basins and to the final stabilization of the craton.

We argue here that the history of the Kaapvaal craton can be understood in terms of processes similar to those that occur today at mid-ocean ridges, ocean subduction zones and continental margins. During the first period of continent formation, intraoceanic processes dominated, resulting in continental nuclei (shields) which may have been similar to the oceanic plateaux now present in the Western Pacific. We argue, moreover, that the first separation of this early continental lithosphere could have occurred only when the mean elevation of mid-ocean ridges had sunk below sea level. During the second period, the continental nuclei amalgamated by means of collisional processes which, together with internal chemical differentiation, created the first stable continental kernels (cratons).

We now give an overview of the Kaapvaal craton to provide a framework within which to discuss its formation.

Regional framework of Kaapvaal craton

The craton covers about $1.2 \times 10^6 \text{ km}^2$ (Fig. 1). Geological mapping and geochronology have largely defined the outer limits at surface, but geophysics and image processing have further helped in locating the craton boundaries^{6,7}. To the north, the craton is bounded by the Zimbabwean craton; to the west and south, it is bordered by Proterozoic mobile belts; and to the east by the Lebombo monocline of Jurassic volcanics associated with the break-up of Gondwana (Fig. 1a). In depth, the craton has also been reasonably well defined by seismologists and mantle petrologists. Waveform inversion studies have demonstrated that there is a strong correlation between the surface expression of the craton and velocity variations from crust to upper mantle^{7,8}. Crustal refraction observations in the Proterozoic mobile belts contrast with the craton refraction. Results suggest that the crust of the Namaqua Proterozoic belt has a substantial (30 km) component of intermediate-velocity rocks ($6.6\text{--}6.9 \text{ km s}^{-1}$) and a depth to Moho of about 42 km. In contrast, the intermediate-velocity rocks in the Kaapvaal craton are represented by only about 15 km of the lower crust, and the depth to the Moho is 37 km (ref. 8).

The upper layers of the craton have low S-wave velocities, with a mean of 4.3 km s^{-1} in the crust and upper mantle down to 80 km depth (about 8% slower than global averages); this is followed by a positive velocity gradient and a higher velocity of about 4.6 km s^{-1} between 110 and 220 km depth⁷. The velocity inversions show that velocities beneath the craton in the depth range 50–350 km are substantially greater than the global average; from this data it can be concluded that the maximum base of the craton is at a depth of about 350 km (ref. 7). This

is consistent with the idea that cratons are associated with high-velocity anomalies world-wide extending to 300–400 km in depth; Jordan⁹ termed this the continental tectosphere.

The craton is characterized by low heat flow ($\sim 45 \text{ mW m}^{-2}$) and the occurrence of diamondiferous kimberlites, whereas the younger surrounding mobile belts have relatively high heat flow ($\sim 80 \text{ mW m}^{-2}$) and barren kimberlites, and thus have lithosphere thicknesses less than 120 km (ref. 10). Geotherms calculated from heat-flow data yield similar information, as does thermobarometry of upper-mantle conditions obtained from kimberlite intrusions¹⁰. Clearly, the craton has a deep, cool and chemically reduced lithospheric root¹¹, comprising about $200\text{--}420 \times 10^6 \text{ km}^3$ of tectosphere, a large proportion of which is Archaean in age¹¹.

At surface, the craton can be subdivided into a number of Archaean subdomains^{12–14} (Fig. 1a). Teleseismic P- and PkP-wave travel-time residuals suggest that some of the subdomains are also well defined at depth, and that local changes in depth to the bottom of the lithosphere correspond closely with subdomain boundaries at surface (E. Peberdy, thesis in preparation). The oldest known subdomains occur in the eastern region of the craton: this is the Ancient Gneiss terrain and the southern Barberton terrain (Fig. 1a). The comagmatic mafic and ultramafic rocks of the southern Barberton terrain represent a remnant of very early oceanic lithosphere (~ 3.5 Gyr) which was obducted, roughly 45 Myr after its formation, onto a volcanic arc-like terrain by processes similar to those which have emplaced modern ophiolites at convergent margins of Phanerozoic continents^{15–20}. This period of thin-skinned thrusting (3.4–3.6 Gyr) in ocean and arc-like environments was followed by amalgamation of the ocean and arc-like terrains at 3.2–3.3 Gyr; the region finally stabilized at ~ 3.1 Gyr (refs 15, 19). Careful consideration of geochronological data from the Barberton and surrounding region thus allows modelling of the first emergence and stabilization of a continental area in excess of $5 \times 10^5 \text{ km}^2$; we refer to this earliest continental region as the Kaapvaal shield, or simply as the shield (Fig. 1, Table 1). Mantle xenoliths from kimberlites indicate that most of the lithospheric mantle of this shield consisted of metamorphosed peridotites and basalts (eclogites)¹¹, and surface geology has shown that the overlying crust was a mixture of mafic and ultramafic (greenstone), and tonalitic-trondhjemitic (granite) rocks, at various metamorphic grades¹². Geochemistry of Archaean and Proterozoic sediments derived from this shield indicates that the near-surface proportion of greenstone to granite must have been roughly 3:7 (ref. 21). Surface exposure of the preserved crust indicates that this ratio is too high (see Figs 1b and 4b). Geophysical observations also support a lower ratio: seismic reflection and refraction data across the central craton, combined with the geological observations, indicate that the Archaean crust-mantle boundary transition was at $\sim 35\text{--}40$ km; and because P-wave velocities exceeding 7 km s^{-1} are only found in the lowermost 10% of this crust²², a substantial upper crust of dominantly felsic composition is evident. A considerable fraction of greenstone material must therefore have been eroded from the upper crustal levels of the craton.

Along the northern margin of the craton, granulite terrains of very different crustal architecture have been tectonically juxtaposed with low-grade granite-greenstone terrains, during two successive late Archaean episodes^{23–25} (Fig. 1a). First, northerly directed thrusting has emplaced the high-grade 'continental-shelf' terrain of the central Limpopo belt and a high-grade granite-greenstone terrain of the northern Limpopo belt on the low-grade granite-greenstone terrains of the Zimbabwean craton^{23–26}. In the southern part of the Limpopo belt, a shield-type granite-greenstone terrain was similarly thrust northward. Single zircon U-Pb analyses indicate that this northward-directed tectonic transportation must have taken place between 2.7 and 2.9 Gyr (ref. 27). At a later stage, southerly directed thrusting, as well as east-west transcurrent shearing, occurred

during the Limpopo orogeny (2.675 ± 0.05 Gyr). These two periods of compressive deformation can be reconciled, respectively, with the development of a North American Cordillera-type accretionary orogeny of composite terranes along the northern margin of the shield^{14,28}, and with later overprinting by an alpine or himalayan-type continent-continent collision^{23–25} (Table 1).

Between the period of the crustal amalgamation (3.2 Gyr) of the shield and the end of the Limpopo orogeny, several important tectonic events affected the central part of the shield. These tectonic processes are reflected in the depositional environments of major overlying Archaean sedimentary basins (Fig. 1).

Mid-Archaean shield development (3.7–3.0 Gyr)

The core region of the Kaapvaal shield consists of at least three well defined tectono-stratigraphic terrains: the Ancient Gneiss terrain, the southern Barberton greenstone terrain and the northern Barberton greenstone terrain (Fig. 1). The boundary zones between these terrains are all tectonic, characterized by thrust and strike-slip faults^{14,15,19,29–31}. The two Barberton greenstone terrains, which together form the Barberton greenstone belt, are part of a fold-and-thrust belt with rocks ranging in age between 3.54 and 3.08 Gyr (refs 14, 15, 19, 20). The Ancient Gneiss terrain to the south contains some older tonalitic rocks (up to 3.65 Gyr)^{32–34}.

Mafic and ultramafic rock sequences of the southern Barberton greenstone terrain, dated at ~ 3.5 Gyr, are similar to parts of Phanerozoic ophiolites¹⁶. But the degree of hydrothermal overprint is considerably greater in this inferred Barberton ophiolite (referred to as the Jamestown ophiolite complex, JOC; ref. 16), and is associated with considerable SiO_2/MgO metasomatism and black-smoker-like mineralization. The hydrothermal overprint occurred during processes analogous to those occurring at mid-ocean ridges, but at greater heat and fluid flux than today^{15,16,30,35,36}. The black-smoker-like mineralization was precipitated from saline brines at temperatures between 90 and 150 °C (refs 15, 37, 38). Oxygen isotope signatures indicate that the H_2O responsible for the precipitation of associated quartz is isotopically indistinguishable from that of modern sea water^{16,37,38}. Fluid inclusion studies suggest shallow marine conditions, possibly no more than 60 m, during this mineralization^{37,38}. Moreover, the postulated hydrothermal discharge vents associated with the mineralization can be traced into banded iron formation, which contains postulated subareal mudpool structures^{38,39}. Sedimentology and geochemistry of associated rocks have also indicated shallow-water environments^{40,41}. These observations could be used to infer (if we apply a modern-day analogy) that the spreading ridges in the Barberton area were very shallow to subareal.

Between 3.46 and 3.43 Gyr, within 50×10^6 years of its formation, the JOC was thrust onto an active arc-like terrain^{15,19,20,42}, with a trondhjemitic-tonalitic basement which dates back to 3.54 Gyr (ref. 20; Fig. 2). Large-scale imbrication of the oceanic-like rocks also occurred before and during its emplacement^{15,17,18,30,42}. A second period of major thrusting (directed northwest) affected all three terrains between 3.3 and 3.2 Gyr (refs 15, 19, 20, 43, 44; Fig. 2). The time of the second thrusting event has been constrained in the central part of the southern Barberton terrain by high-level porphyries (dated at 3.229 Gyr) that are deformed by the thrusting, and similar porphyries (dated at 3.227 Gyr) that cross-cut the thrusts^{15,43}. In the northwestern region of the Barberton terrain this tectonism was accompanied by a second episode of extensive, intermediate, arc-like igneous activity^{15,19,20,44}. The three terrains must have amalgamated by this time because small, chemically similar trondhjemitic intrusions of this age occur in all three terrains^{15–19,43,44}. After amalgamation, transcurrent shearing occurred while regionally extensive granite sheets (the Mpuluzi intrusions and Nelspruit granite) were emplaced between 3.15 and 3.07 Gyr across the three terrain boundaries^{15,44,45}.

Thereafter, radioactive isotopic systems of the supracrustal sequences of this part of the craton were not pervasively disturbed³⁹. Resetting of the Rb/Sr, Sm/Nd and Pb/Pb systems at ~ 2.7 Gyr has, however, been detected in some thin but regionally extensive carbonate horizons and altered volcanics zones, indicating focused fluid activity at elevated temperatures (≥ 200 °C) at that time^{46,47}. A distinct $^{39}\text{Ar}/^{40}\text{Ar}$ mineral plateau at ~ 2.7 Gyr has also been obtained from the central Barberton terrain⁴⁸.

To the south of the Barberton region, greenstone fragments of the southern (Natal) granite-greenstone terrains (Figs 1 and 2) range in age between ~ 3.4 and 3.2 Gyr: the oldest fragments occur in the north and the youngest along the southern margin of the craton^{34,49,50}. The southernmost fragments contain dominantly mafic rocks, but include boninites and komatiites (A. Wilson, personal communication), which have been contaminated with older 'continental' crust⁵⁰. Shallow marine conditions are inferred from studies on associated sedimentary rocks³⁴. Widespread pre- to syntectonic plutonism in this region is of the same age and composition as that of the second episode of intermediate igneous activity affecting the Barberton terrains to the north (3.3–3.2 Gyr), and was similarly followed by post-tectonic granite emplacement before 3.0 Gyr (refs 12, 34, 45). The Natal granite-greenstone terrain and parts of the Ancient Gneiss terrain are unconformably overlain by the relatively undeformed volcano-sedimentary sequence of the Pongola basin (Fig. 1b), the onset of which was before 2.95 Gyr (ref. 34). It is therefore likely that by 3.2 Gyr the Natal terrains had also amalgamated with the Barberton terrains, and that this combined terrain had stabilized before 3.0 Gyr.

Mafic and ultramafic rock sequences along the northern edge of the Kaapvaal shield, in the Pietersburg greenstone belt, resemble the oceanic rocks of the Barberton area and are probably of similar age^{19,51}. The Pietersburg rocks are believed to be part of a much larger allochthonous ophiolitic terrain, which may include similar rocks of the Giyani belt (Fig. 1b) and possibly the high-grade greenstone remnants of the southern marginal zone of the Limpopo belt (Fig. 1)^{25,51}.

In summary, granite-greenstone basement of similar age and composition to those preserved in the Barberton region exists over at least $5 \times 10^5 \text{ km}^2$: it extends to the south in Natal, to the north at least as far as Pietersburg and Giyani, and to the west at least as far as Vredefort (Fig. 1). This estimate is a minimum, because the craton is only a fragment of a craton that was once much larger before the breakup of Gondwana⁵². Evidence from inclusions in diamonds from Cretaceous kimberlites, which traversed the craton like superdeep 'drill-holes', indicates that by the time of amalgamation recorded in the Barberton area (3.2–3.3 Gyr), a lithospheric 'keel' to this region was already at least 150 km thick^{53,54}. From the observations of this granite-greenstone terrain, we suggest that the mechanisms of lithospheric growth in the mid-Archaean era could be similar to those that currently operate along mid-ocean ridges and along oceanic convergent margins. But the interaction between these tectonic processes may require a geodynamic framework slightly different from today's, as suggested by the large volume of the granitoid component of the crust of this shield and the thickness of its lithospheric mantle. We analyse this further below, and present our model for this new framework before we continue to trace the second period of Archaean craton history.

Earliest continental lithosphere

Relative to most Phanerozoic ophiolites, the Barberton JOC sequence is thin ($< 3.5 \text{ km}$)¹⁶, and this implies either that (locally at least) the ~ 3.5 -Gyr oceanic-like crust was thin^{16,54} or that there has been considerable, but as yet undetected, tectonic thinning.

This ancient oceanic-like crust contained a large ultramafic component (25%) which was pervasively ($> 95\%$) hydrated, with H_2O contents ranging between 2 and 15 volume % (refs

16, 35, 36, 54). The assemblage is thus calculated to have had a density as low as $\sim 2.67 \text{ g cm}^{-3}$ (ref. 54). Such intensely hydrated, Mg-rich oceanic crust and upper mantle, as represented by the JOC, could thus have resisted a recycling (subduction-like) process, leading instead to obduction-dominated tectonics⁵⁴. The earliest continental nuclei may therefore have originated through regional intraoceanic obduction, giving rise to stacking, tectonic loading and subsidence of the hydrated oceanic thrust stacks^{17,18,30,36,42,54} (Fig. 3a, upper panel). It is postulated that stacking of ocean slabs would have resulted in melting of the lower parts of the thrust to yield extensive trondhjemite-tonalite melts under variable pressure, temperature and H_2O conditions at depths of 20 km or more^{42,55} (Fig. 3a, lower panel). Thus, the 'essential ingredient' needed to drive this obduction-dominated tectonism and granitoid formation is a buoyant and hydrated oceanic crust: an elevated MgO and H_2O content would provide this in the form of the low-density mineral serpentine, which dominates the mineralogy of the Barberton ophiolite¹⁶.

It is generally accepted that the Archaean oceanic crust had a much higher MgO content than oceanic crust today^{56,57}. Such a crust, however, would have remained relatively dense unless H_2O had efficient low-temperature access to the MgO-bearing minerals (such as olivine and pyroxene) to convert them into low-density serpentine minerals. Furthermore, H_2O must be added to the oceanic crust to facilitate the formation of granitoid magmas⁵⁸. The most efficient process known for hydration of mantle material takes place at mid-ocean ridges (MOR) during the formation of oceanic lithosphere. We speculate, therefore, that before 4.0 Gyr the average MOR stood above sea level, so that 'dry recycling' of the Earth's oceanic lithosphere prevailed, and therefore little continental crust formed (Fig. 3b). Such a model is compatible with the geochemical observations that the earliest preserved mafic-ultramafic rocks on Earth were derived from mantle sources that already had a long history of relative chemical depletion and extraction of even earlier, but recycled oceanic crust⁵⁹⁻⁶¹. As a corollary, in our model the formation of the earliest continental regions was related to a time that the Earth's MORs 'drowned' below sea level, triggering the first efficient and continuous interaction of our planet's hydrosphere and mantle. Although this may have been a gradual process, it nevertheless must have been an important revolutionary episode in the history of Earth's differentiation. Because of the Earth's decreasing internal heat production, with time the average oceanic crust has gradually evolved to a more tholeiitic composition, with less Mg and more Fe than that of the Archaean oceanic crust⁵⁷. Concomitantly, densities of the oceanic crust must have increased⁵⁷. We believe that this would have gradually heralded the return from oceanic obduction to subduction (Fig. 3b).

The initial rate at which the differentiation (continental crust formation) may have taken place is not known, but in our model this must have been related to prevailing rates of oceanic lithosphere formation, which is believed to have been up to 2-3 times as great as today⁵⁷. The immediate products of these hydrosphere-mantle interactions yielded granite-greenstone nuclei such as the Kaapvaal shield, which, together with their depleted dunitic-harzburgitic keels (Fig. 3), provided continental kernels for further (plate) tectonic accretion and collisions.

If our interpretations are correct, then the occurrence of similar rocks in the wider craton framework implies that the later (3.2-3.3 Gyr) processes of amalgamation of granite-greenstone terrains, and further internal melting of the nucleus to yield the late high-level granites (crucial for transporting radioactive heat-producing and other large-ion lithophile elements from the shield's lower crust to its upper crust, as documented in detail in refs 45 and 62), all had an important part in the final consolidation of the shield by the end of mid-Archaean times. By 3.1 Gyr the shield was firmly established and constituted a framework that could support a series of late Archaean sedimentary basins (for example the Witwatersrand, Pongola and Ventersdorp basins) which have all the characteristics of a

wide variety of modern sedimentary basins (see below). In the phraseology of geologists, geophysicists and geochemists, therefore, the Kaapvaal shield was a true continent (Table 1).

Late Archaean craton development (3.1-2.6 Gyr)

The quest to decipher further the continued evolution of the Kaapvaal shield into a stable cratonic region is aided, in southern Africa, by a vast and thick succession of late Archaean sediments that overlie the rocks of the shield. These sediments have preserved a rich chronological record of the processes of further continental growth. Advances in 'reading' this stratigraphic record, particularly using U-Pb zircon studies, have recently been made. Both in the centre and along the peripheries of the craton, these sediments can be observed in direct contact with their underlying basement; here, any late Archaean tectono-thermal processes that have affected the shield rocks can be equated with those preserved in the late Archaean sedimentary successions. We first review these data below and then use the information to model the consolidation of the craton.

Sedimentary basins. Following the event of crustal amalgamation at ~ 3.2 Gyr and further lithospheric stabilization documented in the southern sectors of the craton, the shield area suffered a period of extension reflected by extensive volcanics at the base of regional sedimentary basins which covered large sectors of the shield^{12,63-67} (Fig. 1a and b; Table 1).

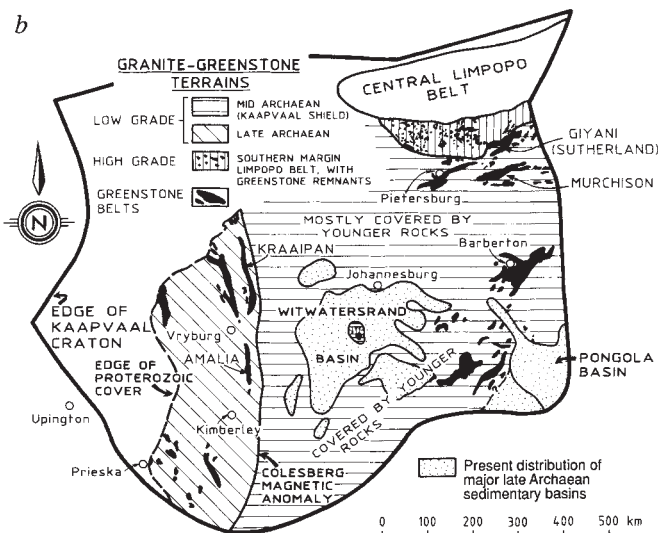
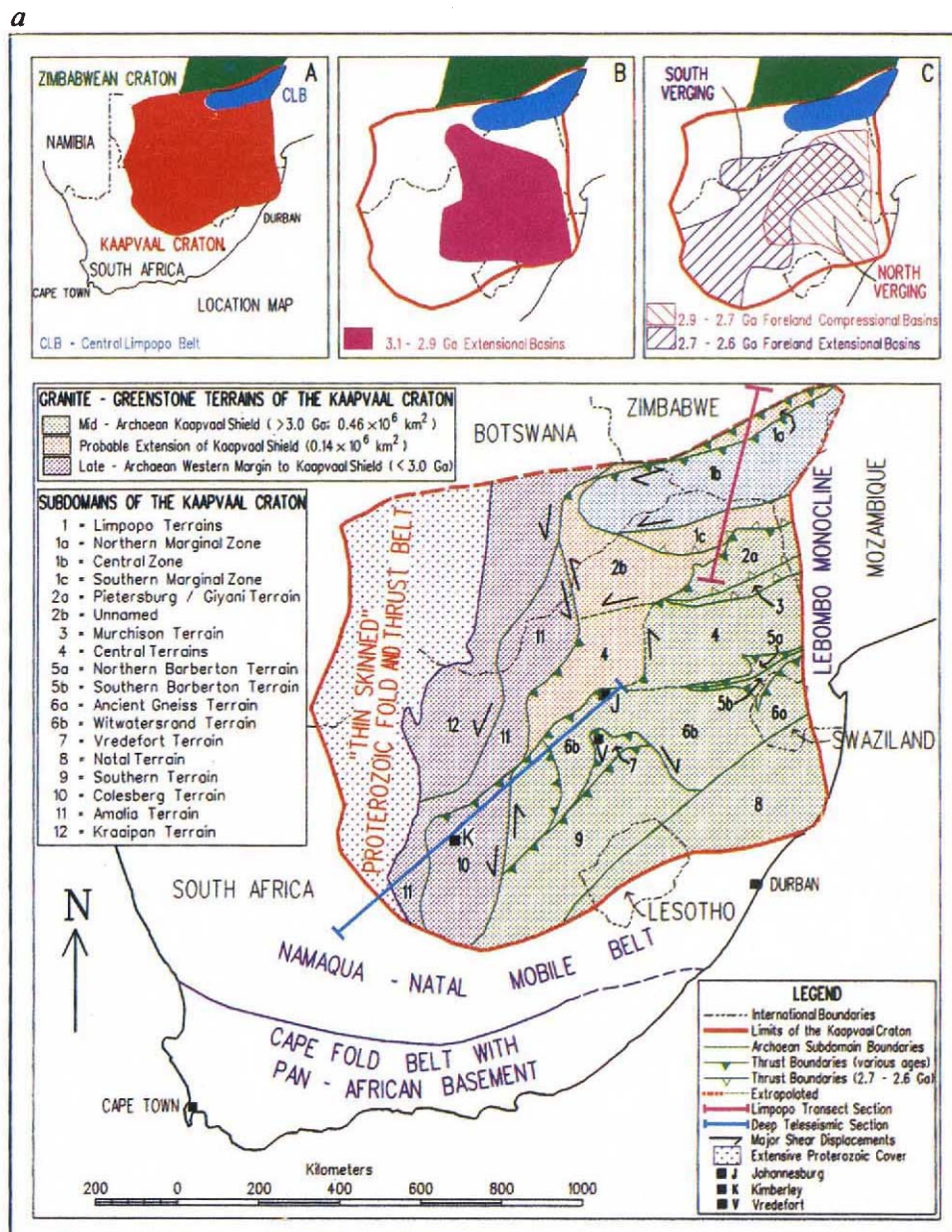
The Dominion Group volcanics heralded the start of the Witwatersrand sedimentary basin^{12,67-69}. Stratigraphic studies indicate that the Witwatersrand basin is a polyphase basin composed of at least two cycles^{63,67}. The lower part of the Witwatersrand basin probably formed on a continental platform of a passive continental margin, facing open-ocean conditions to the south and the west (W. E. L. Minter, personal communication). Zircon geochronology indicates that the basin must have been initiated between 3.07 and 3.12 Gyr and supplied with sediments from a predominantly northern source ranging in age between 3.07 and 3.35 Gyr (refs 67, 68). This lower part of the Witwatersrand group is of the same age as the post-rift, thermal subsidence of the Pongola basin along the southern margin of the craton^{69,70} (Fig. 1b). Stratigraphy and isotope work on Pongola carbonates indicate that the southernmost outcrops of this basin were also part of a passive continental margin facing open ocean conditions to the south (ref. 70 and P. E. Matthews, personal communication).

The upper part (second cycle) of the Witwatersrand basin formed in an environment influenced by convergent tectonic activity¹² and probably represents a polyphase (successor) foreland basin type^{63,64}, formed between 2.71 and 2.84 Gyr (ref. 67 Fig. 1a). This was followed by another episode of major extensional tectonics expressed by the Ventersdorp volcanism and sedimentation, now known to have formed between 2.7 and 2.6 Gyr (ref. 67; Fig. 1a).

Because the depositional environment of the Witwatersrand basin is in part reflected in the thermotectonic history of the underlying basement rocks, now exposed in the Johannesburg dome along the northern edge of the basin, and the Vredefort structure in the centre of the basin (Fig. 1), we first summarize the relevant geological events in these areas.

Vredefort structure. The Vredefort structure, which is thought to be the oldest and the largest known impact crater on Earth^{71,72} (albeit tectonically modified)⁷³, is an ideal locality in which to study the basement of the Witwatersrand basin. The structure consists of an uplifted core of Archaean crystalline rocks surrounded by a collar of Witwatersrand strata (Fig. 4). Since the postulated impact at 2.0 Gyr the structure has been tectonically tilted⁷³, and in the exposed northwestern parts of the structure, the Archaean crystalline rocks and the Precambrian strata are now nearly vertical in attitude (Fig. 4a). A northwest-southwest traverse across the structure indicates that a ~ 36 -km section, across the entire Archaean crust and the overlying strata (Fig. 4b), may be exposed on surface⁶².

FIG. 1 a, Schematic diagram (based on geological, aeromagnetic and gravity data) showing the Archaean crustal subdomains of the Kaapvaal craton referred to in the text. At present, the geological data do not let us distinguish whether these subdomains are allochthonous with respect to each other (and thus significantly displaced from their original sites of formation), or were once part of a larger coherent geological province that has subsequently been tectonically disrupted. It is therefore unclear whether the domains should be classed as terranes (allochthonous; see ref. 5). To emphasize this lack of knowledge, we refer to the subdomains as terrains. Inset A shows the locations of the Kaapvaal craton, the central zone of the Limpopo belt, and part of the Zimbabwean craton. Inset B schematically shows the extent of the late Archaean extensional sedimentary basins (3.1–2.9 Gyr). Inset C schematically shows the extent of the subsequent late Archaean foreland basins formed during extended periods of compression across the craton, culminating in the 2.7 Gyr collision between the Zimbabwean and Kaapvaal cratons. The red line (Limpopo transect section) locates the section shown in Fig. 5. The blue line shows the deep teleseismic line across the southwestern part of the Kaapvaal craton along which P-wave experiments are being conducted (E. P., thesis in preparation). The minimum depth to the bottom of the lithosphere beneath the central craton varies from 170 to 310 km; distinct changes in depth occur directly beneath the different terrain boundaries delineated at surface (E.P., thesis in preparation). b, Distribution of the granite-greenstone terrains of the Kaapvaal craton from surface geology¹² and aeromagnetics⁶. Note the change in regional tectonic anisotropy as delineated by the greenstone belts, from northeast-southwest in the older central and southern part of the craton, to north-south in the younger western parts of the craton. Note also that in two regions (the southern margin of the Limpopo belt and within the Vredefort structure), the granite-greenstone terrains are at granulite-grade of metamorphism, exposing deep crustal sections of the Archaean cratonic crust (see also Figs 4 and 5). Also shown are the present-day extents (from geology and geophysics) of the Witwatersrand and Pongola basins. There is general consensus that these basins are erosional remnants of originally much larger sedimentary depositories, as shown schematically on Fig. 1, inset B (personal communications with participants of the Precambrian Sedimentary Basins conference, Pretoria, 1991).



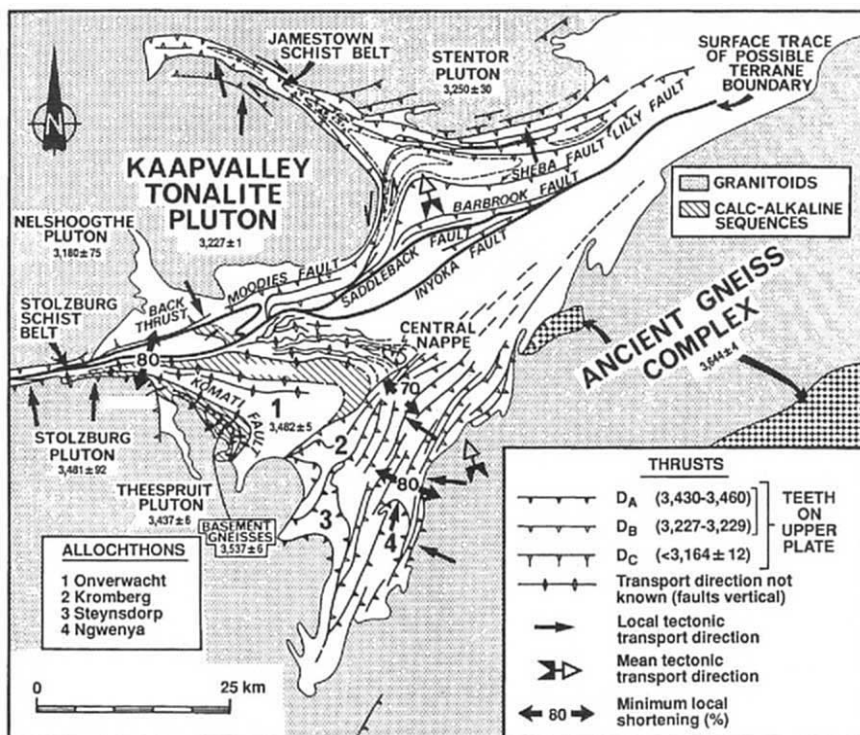


FIG. 2 Principal thrusts, strike-slip faults and allochthons now recognised in the Barberton fold and thrust belt^{15,19}. A large structural break along the Inyoka-Saddleback fault system separates the different lower lithostratigraphic sequences of the northern and southern Barberton terrains of Barberton¹⁹. Gravity data show the belt to be ~4 km thick, with some 'roots' having a maximum thickness of 8 km (ref. 13); note that old 'continental' basement rocks occur as tectonic slices within the greenstone belt. Ages given (in Myr) refer only to thrusts shown in central zone of the belt, immediately south of the surface trace of the possible terrain boundary; many of the northeast-southwest thrust faults may in fact be of D₂ age (ref. 15).

The evidence offered by the Vredefort profile indicates that tectonic stacking of thick (>10 km) continental lithosphere slabs played an important role in the evolution of the craton. In particular, the transition from upper-crustal amphibolites to lower-crustal granulites is characterized by a brittle-ductile shear zone known as the Vredefort discontinuity^{62,73}. Structural and geochemical data indicate that between 2.5 and 2.8 Gyr, the amphibolitic upper granite crust (outer gneiss granite, or OGG) was tectonically juxtaposed above granite-greenstone crust (Inlandsee leucogranofels, ILG) along an extensive thrust plane located ~8 km beneath the base of the Witwatersrand basin (Fig. 4b). Vibroseis data has shown that this thrust plane is of regional extent⁷⁴.

The lowermost rocks of the steeply dipping section, ~36 km into the section, consist of serpentized harzburgites⁷⁵. These harzburgites are thought to represent uppermost-mantle rocks of the craton; geochemical evidence suggests that these rocks were once part of Archaean oceanic lithosphere and may be equivalent to parts of the ~3.5-Gyr JOC exposed in the Barberton terrain⁷⁵. The harzburgites were tectonically emplaced beneath the granitoid rocks and metamorphosed to ultramafic granulites, consistent with a model of Archaean tectonic, as opposed to igneous, thickening of the shield^{17,18,30,54}.

Johannesburg dome and northern greenstone belts. Principal north-verging thrust faults along the southern edge of the Johannesburg dome interleave granite-greenstone basement with Witwatersrand sediments^{76,77}. This tectonism can be placed within the time bracket of 2.4–2.8 Gyr, that is, within the range of the thrusting recorded at the Vredefort structure. Similarly, field work and zircon dating in the Pietersburg and Giyani greenstone belts, situated along the northernmost edge of the low-grade granite-greenstone terrains of the craton (Fig. 1), has shown that major northward thrusting along the northern margin of the craton occurred between 2.7 and 2.9 Gyr (refs 19, 27, 51, 78). Coarse clastic terrestrial foreland basins, preserved in the Pietersburg and Giyani greenstone belts, formed syntectonically with this northward-thrusting, riding 'piggy-back' on the top of main thrust zones^{19,51}. These basins are therefore age-equivalents of the upper Witwatersrand basin and, like the upper Witwatersrand basin, their sediments were derived predominantly from the north^{19,27,51}.

In summary, there seems to be a regional tectono-stratigraphic link between the shield's granite-greenstone terrains from at least as far south as the Vredefort terrain to the edge of the Limpopo terrains, through a regional north-verging foreland basin (or series of intermontane 'piggy-back' basins). How do these events relate to the orogenic activities of the Limpopo belt along the northern margin of this shield?

Limpopo belt. The Limpopo belt provides a complete section from the edge of low-grade granite-greenstone terrain of the shield to the high-grade granulite terrain and then to the low-grade granite-greenstone terrain of the Zimbabwean craton^{23–26} (Fig. 1). The granulite terrain has been interpreted as the root zone of an orogenic belt formed during the collision of the two cratons in late Archaean times^{12,23,26}. The belt can be subdivided into three main subterrains: the southern marginal, the central and the northern marginal zones (Figs 1 and 5). Geological and Pb-isotope data indicate that the central zone constitutes a distinctly different terrain from those to the north and south, and is separated from them by tectonic breaks^{23,25,79}. Data on seismic reflection, magnetic, gravity and electrical conductivity all support models that incorporate major northward-thrusting across the entire belt (Fig. 5)^{13,24,25}; the central zone has also been displaced westward⁷⁹ (see also Fig. 1a).

These northward-directed events were followed by crustal shortening and considerable shear-zone development in the form of south-verging thrusts and northeast-southwest striking strike-slip faults^{23–25,78} (Fig. 1a and Fig. 5). The shear-zones were directly responsible for the rehydration of overthrust granulite-facies assemblages; away from these zones the granulites preserve classic decompression textures^{23,24}. The central and southern marginal zones indicate rapid uplifts of ~20 km at 2.675 ± 0.05 Gyr (refs 23, 24). The geological data suggest that this uplift was due to the formation of a classic tectonic 'pop-up' structure²⁴, and that the southward thrusting can be interpreted as a strong back-thrust^{24,78}; that is, part of retrocharriage such as documented in many Phanerozoic orogenic belts⁸⁰. Retrograde granulites of the southern marginal zone occur as isolated nappes (klippen) above prograde granite-greenstone rocks up to at least 50 km south of the thrust front²⁴. Beneath the Limpopo orogenic belt, the present crust thickness is similar to thicknesses beneath Palaeozoic mountain chains (30–35 km), but it thickens

northwards to more than 40 km beneath the Zimbabwean cratonic foreland^{24,81}. Along the southern margin there is an apparent 5-km offset in Moho depth across the major south-verging backthrust (Fig. 5). Vibroseis reflection data across the belt show a clear fanning pattern of crustal reflectors²⁴ similar to the characteristic fanning crustal-reflection patterns of many Phanerozoic and some Proterozoic orogens, including the Alps, Pyrenees, Appalachians, Caledonides, Svecofennides and Karelides⁸². In each of these orogens, the fanning reflection within the crust seems to mark tectonically imbricated crust on opposite sides of a suture zone. By analogy, we interpret the offset Moho beneath the southern marginal zone of the Limpopo belt as the site of underthrusting of the continental crust of the shield within the collision zone.

West of the Colesburg lineament. Little is known about the terrains along the western margin of the shield, to the west of the Colesburg magnetic lineament (Fig. 1b). Age determinations on detritus derived from these terrains as well as sporadic basement outcrops indicate late Archaean ages (2.5–3.0 Gyr)^{68,69}. It has been suggested that these terrains represent the roots of calc-alkaline arcs⁶⁴. A north-south tectonic 'grain' in this western margin is clear from the geology⁸³ as well as from aeromagnetism^{6,84} (Fig. 1b). When linked with Limpopo-aged north-south shortening over the whole craton, left-lateral wrench movements must have operated along the western margin of the craton: the terrains along this margin may therefore be equivalents of the younger (2.6–2.7 Gyr) north-south trending arc and back-arc-basin terrains of western Zimbabwe⁸⁵, and

could be (displaced) fragments of an extensive late Archaean, north-trending convergent boundary along the western margin of the newly amalgamated southern African shield (Kapaal-Zimbabwean cratons combined). The tectonics along this western margin also affected sedimentation along the western margin of the uppermost Witwatersrand basin⁶⁴.

Consolidation of an Archaean continent

The inferred time of juxtaposition of the lithotectonic terrains at lower crustal levels documented at Vredefort, and the upper-crustal northward thrusting in the Witwatersrand basin is within the age range of the extensive northward thrusting that occurred between ~2.7 and 2.9 Gyr along the northern margin of the craton described above. The available data suggest that during this period, much of this northern margin may have been assembled through a major amalgamation of elongate terrains (some of which may have been 'exotic'): examples are the Murchison belt (which is an arc-like terrain⁸⁶), the central terrain of the Limpopo mobile belt (which is a continental fragment with associated shelf sediments⁸⁷), and the Pietersburg, Giyani and southern Limpopo Marginal terrains (as indicated previously, these are probably composites of the early granite-greenstone terrains that constitute the shield; Fig. 1)^{19,24,51}.

Assembly and amalgamation of the terrains along the northern margin in the central region, and along the western margin of the craton, all overlapped with the period of sedimentation in the upper Witwatersrand basin and in the northern greenstone belts (bracketed by the ~2.7 and 2.9 Gyr deformation events

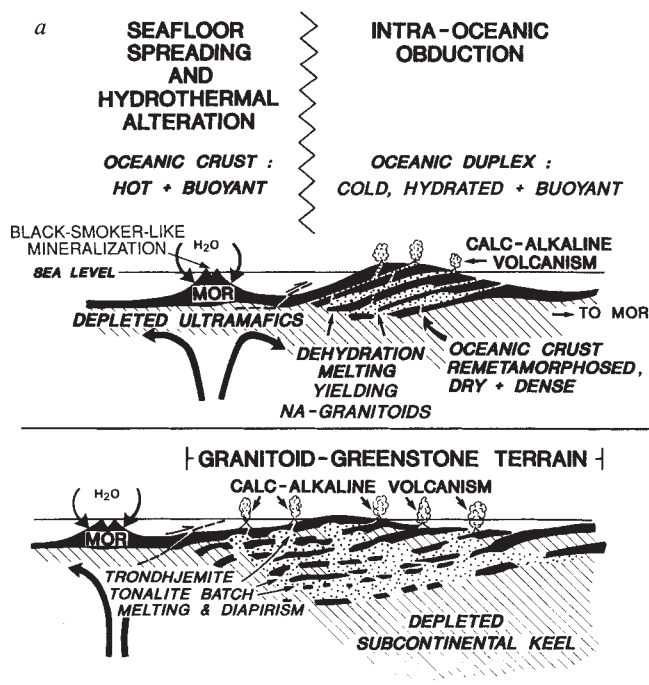
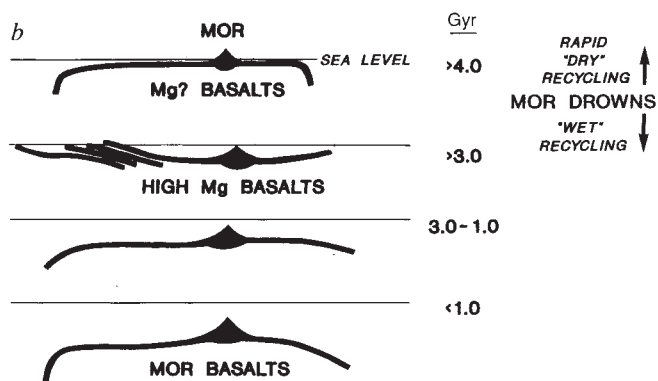


FIG. 3 a, Schematic model depicting the formation of the Earth's earliest preserved continental lithosphere, through linked processes of spreading and obduction-dominated tectonics. Heat flux and hydrothermal interaction of Archaean MORs was more intense than today^{16,30,35,36} causing the formation of chemically enriched MgO crust (because of hotter mantle melts^{56,57}), but physically less-dense oceanic lithosphere (because of associated greater degrees of hydration, and silica, iron and magnesium metasomatism^{16,30,35,36}), than currently. Consequently, continuous imbrication of oceanic lithosphere led to the formation of intraoceanic thrust-stacks (upper panel). This resulted in regional load-induced subsidence (lower panel), which in turn caused *in situ* differentiation and density inversions during dehydration melting of the 'wet' oceanic-like rocks to form the vast amounts of tonalite-trondhjemites and associated calc-alkaline extrusives now present in and around greenstone belts such as those preserved in the Barberton



terrains. These intraoceanic processes produced a shield area consisting of a mafic-ultramafic keel and an overlying continental-like crust by tectonic rather than igneous thickening. b, Schematic model for the evolution of oceanic plates with respect to sea level, and related formation of the earliest continental crust (granite-greenstone terrains) and lithospheric keel (peridotite-eclogite) of the Kaapvaal craton, following gradual 'drowning' of MORs at ~4.0–3.6 Gyr. Submergence of MORs was probably related to both the decreasing heat content of the Earth and the ageing of average oceanic lithosphere. It therefore reflects a thermally controlled increase in the average depth of the ocean basins^{57,95}. No interaction between the hydrosphere and mantle occurred before this, so that 'dry' recycling of oceanic lithosphere prevailed; consequently continent formation was inefficient or absent⁵⁸. Gradual decline in the magnesium content of the oceanic crust (caused by declining average mantle temperatures⁵⁷, and thus the decrease of the volume of low-density serpentinite, heralds the end of intraoceanic obduction-dominated tectonics and the start of modern-day subduction processes of tholeiitic oceanic crust. Although the detailed time framework for this transition is not known, Phanerozoic-like ophiolites of ~2.0 Gyr in age⁸² have now been recorded, suggesting that by the early Proterozoic, oceanic processes were similar to those of today, but with a greater potential for shallow-angle subduction⁹⁵, as schematically shown in the third frame of b. Studies on Proterozoic ophiolites and associated sediments^{82,100–102} (and P.E. Matthews, personal communication), however, indicate that modern-type MOR-basalts chemistry, and ophiolite thicknesses comparable to those of the Phanerozoic, as well as probable MOR-depths similar to the present day, were attained between 1.0 and 2.0 Gyr.

and the deposition of the Ventersdorp lavas dated at 2.714 Gyr⁶⁷. One of the chief problems is the observation that the northward thrusting recorded in the Witwatersrand basin (and its underlying basement) coincides with a consistently northerly and westerly derivation of the sediments⁸⁸ of the basin before the Limpopo orogeny. Clearly, the source regions for most of the sedimentation of the Witwatersrand basin cannot be directly related to the Limpopo 'highlands', and the basin is therefore not a simple foreland basin to the Limpopo belt as previously suggested^{63,89}. One way to account for these observations is to model the events in two stages as follows.

Primary orogenic collage. Early (2.7–2.9 Gyr) northwards thrusting during an extended period of tectonic accretion of oceanic and continental fragments formed a wide orogenic belt along the northern margin of the shield^{14,27,28}. This could be compared with the Cenozoic accretionary orogeny bounding the Canadian

shield in British Columbia and southern Alaska^{5,90}. In this model, subduction was to the south, and the accretion process and the subduction zone periodically shifted northwards. During this tectonism, sediments (containing detrital gold and diamonds) were shed southward into orogenic (piggy-back) basins which were overprinted by out-of-sequence northward thrusting. In such a scenario the upper Witwatersrand basin(s) and the northern greenstone-belt basins might be comparable with the (gold-bearing) intermontane basins within, and north of, the accretionary orogeny of southern Alaska.

Collision orogeny and associated extrusion, extension and collapse. Late (2.68 Gyr) southward thrusting documents the final collision of a much larger continental fragment (the Zimbabwean craton) causing backthrusting (retrocharriage) along the southern edge of the Limpopo belt and the formation of a classic 'pop-up' structure of the central Limpopo orogeny (Fig. 5), which can

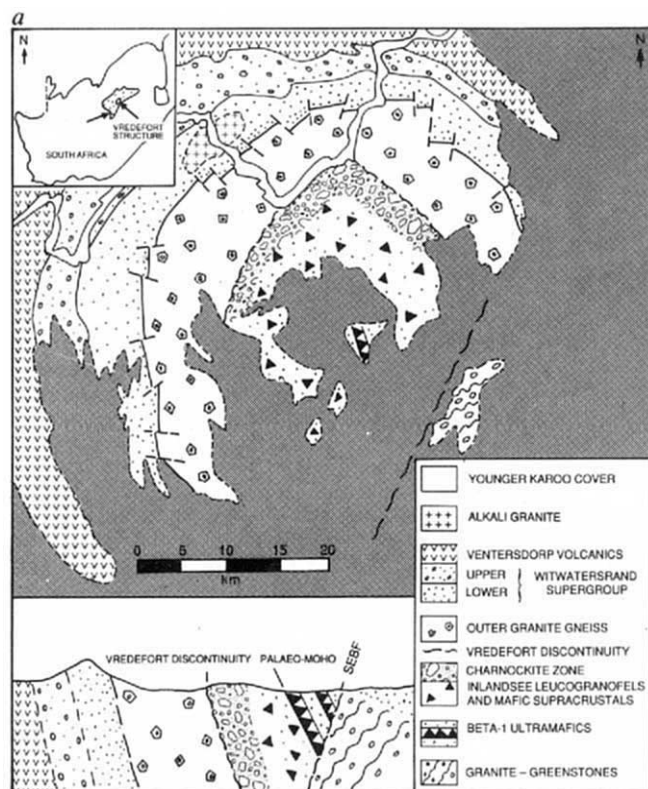
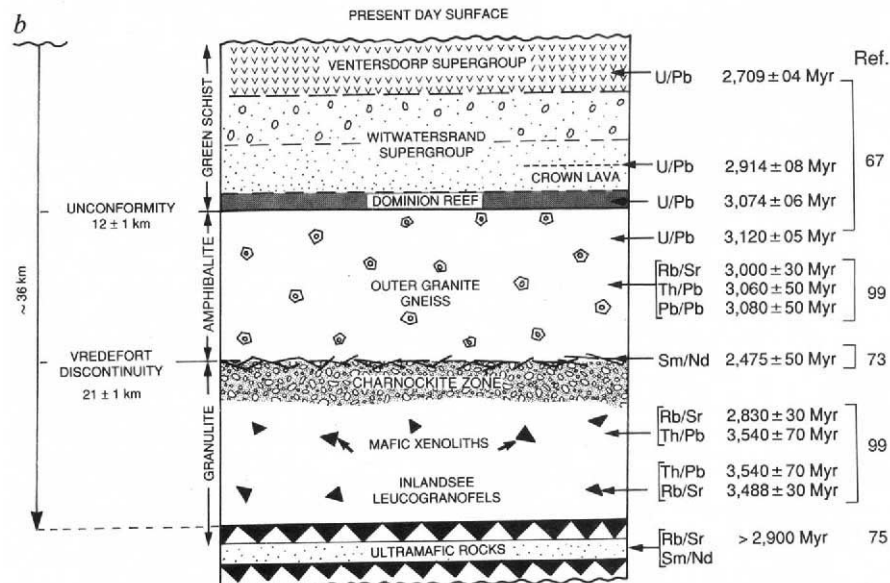


FIG. 4 a, Geological map and a schematic cross-section of the Vredefort structure (modified from refs 62, 73, 75; SEBF, southeast boundary fault). At least four geological domains have been identified in the basement core of the Vredefort structure. Their age relationships and stratigraphic positions in the crust, before uplift 2.0 Gyr ago, are shown in b. Following the section from the northwest to the southeast, the upper 10 km (in contact with the overlying Witwatersrand strata) consist of a sequence of granitic and granodioritic rocks in amphibolite facies, termed the outer granite gneiss (OGG) terrain. Beneath the OGG, a heterogeneous granulite facies domain consisting of charnockites-enderbites, granulites (mafic and felsic) and supracrustal rocks is exposed. Collectively this lower crustal sequence is known as the Inlandsee leucogranofels (ILG) terrain. The OGG and the ILG are separated by a brittle-ductile shear zone known as the Vredefort discontinuity. The deepest rocks exposed near the core of the Vredefort structure consist of Archaean harzburgites which are thought to represent the upper mantle of the Kaapvaal craton. To the southeast of the ultramafic rocks, the structure is transected by the northeast-southwest-trending SEBF. Field evidence indicates that the age of SEBF is earlier than Witwatersrand. The basement to the southeast of the SEBF is a granite-greenstone domain at greenschist facies. The three different terrains (the OGG, the ILG and the low-grade granite-greenstone terrain) were probably tectonically juxtaposed between 2.5 and 2.8 Gyr. b, Composite stratigraphic section, with representative isotope signatures, through the crust into the uppermost Archaean mantle of the Kaapvaal craton, as represented by the Vredefort 'crust on edge' profile.



adequately account for both the regional decompressional metamorphic reactions of the Limpopo orogeny²³⁻²⁵ and the regionally inverted metamorphic gradients along the northern margin of the craton⁷⁸. This south-directed overthrusting juxtaposed granulites of the central Limpopo orogeny across the granite-greenstone terrain to the south; as mentioned before, granulite klippen are at present exposed up to 50 km south of the surface trace of the main south-verging thrust²⁴ (Fig. 5). This backthrusting was important across the whole craton: regional southward thrusting has been recorded as far south as the Murchison terrain⁸⁶, and the shield may have been affected as far south as Barberton, where rocks were disturbed at 2.7 Gyr, as recorded by the low-temperature fluid activity and resetting of the Ar-Ar, Rb-Sr, Pb-Pb and Sm-Nd isotope systems. This isotope resetting may well be related to regional tectonic-induced fluid transportation away from the Limpopo orogenic front, as has been shown for regional fluid migration through extensive sections of the crust ahead of Phanerozoic collision zones during final consolidation of major continents such as the Pangea^{91,92}. Considerable crustal extension and associated crustal remelting in the form of late granite-syenite intrusions occurred throughout the craton during this period^{12,44,45}, as did the eruptive sequences within the extensive Ventersdorp supergroup^{12,63-65,67} (Fig. 1c). This probably represented a period of orogenic extensional collapse^{80,93}. Similarly, in this model, the westward displacement of the central zone of the Limpopo belt is probably a manifestation of extrusion tectonics (compare with refs 80 and 93) between the two colliding cratons.

The above processes mark the end of an extended period of Archaean consolidation of a large section of the shield into a true craton (that is, a stable continent).

Continental lithosphere growth

The Kaapvaal craton was established over a period of at least 1 Gyr. The first $\sim 5 \times 10^8$ years (3.7-3.2 Gyr) saw the emergence of a 'continental' shield area through large-scale intra-oceanic thrusting, tectonic burial and melting of hydrated oceanic crust. The establishment of this continental platform and its underlying tectosphere thus proceeded by tectono-igneous differentiation, driven essentially by the addition of water to mafic-ultramafic crust. Considering the large ratio of granitoid to greenstone in the shield²¹, and taking into account the knowledge that the

granitoids were derived from the greenstone equivalents (hydrated basalts of same age and geochemistry)^{42,45}, it is clear that ocean-floor recycling and associated tectonic processes (subduction/obduction) must have operated.

It is possible that modern counterparts of shield are to be found amongst the present-day oceanic plateaux of the Western Pacific, such as the Ontong-Java Plateau (OJP). Like the Pacific ocean plateaux, which are destined to collide with oceanic and continental arcs^{5,90,94}, the Kaapvaal shield apparently collided with arcs and oceanic terrains. It is therefore interesting to note that the OJP is similar in size to the Kaapvaal shield⁹⁴ and that the OJP is now colliding with, and obducting across, the Solomon Islands arc/back-arc terrain⁹⁴, which has already been specifically compared to Archaean greenstone belts⁹⁵. The OJP has a similar crust thickness and similar seismic velocities to those postulated for the shield (35-40 km; 6-7 km s⁻¹; refs 22, 94). Kimberlite-like intrusions (alnöites) into the OJP contain deep-seated xenoliths (garnet lherzolites) indicating a thick (up to 135 km) depleted oceanic lithosphere beneath the plateau⁹⁶. The depleted mantle beneath the OJP has been underthrust by hydrated oceanic crust⁹⁶, and the ocean plateaux in general are believed to contain high-Mg basalts and komatiites⁹⁷. But a problem that must be considered when comparing the Kaapvaal shield to present-day ocean plateaux is that, according to our model, plateaux-like OJP should also contain large volumes of tonalites.

In the case of the Kaapvaal shield, intra-crustal melting yielded an uppermost granite 'layer' during the period of amalgamation, converting the shield into a truly continental platform of at least 5×10^5 km², with 40-km-thick crust and a lithosphere boundary between 170 and 350 km in depth by 3.1 Gyr. The lithospheric growth rate of the shield, given its formation between 3.2 and 3.6 Gyr, must therefore be ~ 0.5 km³ yr⁻¹; and the crustal growth rate must be ~ 0.05 km³ yr⁻¹ (as opposed to the global Phanerozoic crustal growth rate of 1.06 km³ yr⁻¹ calculated by Reymer and Schubert⁹⁸). If we assume that the total crustal growth rates in the Archaean were 3-4 times the present Phanerozoic rate^{1,37,95,98}, then between 3.2 and 3.6 Gyr, the world's continental surface area must have increased by 64-85 times the size of the Kaapvaal shield (~ 15 -20% of the present continental surface area). As there are at most 10 preserved Archaean cratons worldwide (of which only two substantial

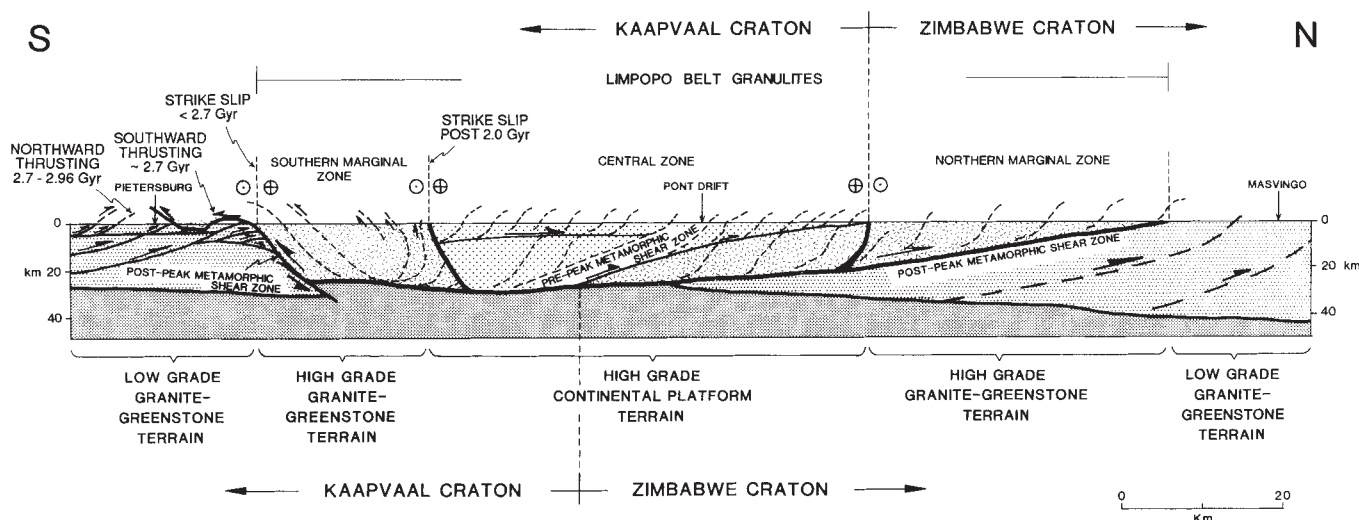


FIG. 5 A composite north-south profile across the Limpopo belt (see red line on Fig. 1a for line of section) based on geological and geoelectric, magnetic, seismic (refraction and vibroseis reflection) and gravity data (compiled from refs 24-26, 78). This profile is similar to those across several Phanerozoic and Proterozoic mobile belts⁸². The profile clearly shows the early northward-directed tectonic transport (dated 2.7-2.9 Gyr) overprinted

by the ~ 2.7 Gyr southward-directed thrusting related to the climax of the Limpopo orogeny. The 2.7 Gyr tectonic loading of the Kaapvaal craton by the Limpopo terrain probably induced focused fluid migration throughout the cratonic crust, at least as far south as the Barberton terrains. Dark shading represents upper mantle. $\odot$, transcurrent movement out of the page; $\oplus$, transcurrent movement into the page.

ones are older than 3.0 Gyr), it is difficult to avoid the conclusion that Archaean continental crust was recycled back into the mantle. The implications of this will be discussed elsewhere (M.J.d.W. and W. Wilsher, manuscript in preparation).

During the second 5×10^8 -yr period (3.1–2.6 Gyr), the shield extended by a further $\sim 5 \times 10^5$ km². It is, however, not clear at this stage how much of this continental growth was due to a net addition of material newly derived from the mantle.

Finally, the geological manifestations of the tectonic and thermal processes that converted the shield into the craton seem to be identical to those occurring during the Phanerozoic. The kinematic details of the tectonic processes remain less well understood, because the time spans within which Archaean geological processes can be resolved are at least an order of magnitude greater than for Phanerozoic orogenic

processes; and considering the complexities of the latter, this leaves a multitude of tectonic models that still need testing before the formation and growth of this Archaean continent are understood. □

Maarten J. de Wit, Cornel E. J. de Ronde and Marian Tredoux are at the Department of Geology, University of Cape Town, Rondebosch 7700, South Africa. Chris Roering is at the Department of Geology, Rand Afrikaans University, Johannesburg 2050, South Africa. Rodger J. Hart is at the Schonland Research Centre, University of the Witwatersrand, Johannesburg 2000, South Africa (seconded from South African Geological Survey). Richard A. Armstrong is at the Department of Geochemistry, University of Cape Town, Rondebosch 7700, South Africa. Rod W. E. Green and Ellie Peberdy are at BPI Geophysics, University of the Witwatersrand, Johannesburg 2050, South Africa. Roger A. Hart is at the College of Oceanography, Oregon State University, Corvallis, Oregon 97331, USA.

- Ashwal, L. D. (ed.) *Tectonophysics spec. Issue* **161**, 143–352 (1990).
- Bowering, S. A., Williams, I. S. & Compston, W. *Geology* **17**, 969–1064 (1989).
- Compston, W. & Pidgeon, R. T. *Nature* **321**, 766–769 (1986).
- Taylor, S. R. *Tectonophysics* **161**, 147–156 (1989).
- Ben-Avraham, Z. *The Evolution of the Pacific Ocean Margins* (Oxford Univ. Press, 1989).
- Corner, B. *Precambrian Sedimentary Basins of Southern Africa* (Terra abstr.) (ed. Eriksson, P. G.) 5 (Blackwell, Oxford, 1991).
- Chichowich, A. & Green, R. W. E. *Phys. Earth planet. Int.* (in the press).
- Green, R. W. E. & Durrheim, R. J. *J. geophys. Res.* **95**, 19927–19932 (1990).
- Jordan, T. H. *Phil. Trans. R. Soc. A* **301**, 359–373 (1981).
- Jones, M. J. *J. geophys. Res.* **93**, 3234–3260 (1988).
- Gurney, J. S. *Afr. J. Geol.* **93**, 423–437 (1990).
- Tankard, A. J. et al. *Crustal Evolution of Southern Africa* (Springer, New York, 1982).
- de Beer, J. H. & Stettler, E. H. *J. Petrol. Spec. Vol.* 163–184 (1988).
- de Wit, M. J. & Roering, C. *The Limpopo Belt: A Field Workshop on Granulites and Deep Crustal Tectonics*, extended abstr. (ed. Barton, J. M.) 42–52 (Rand Afrikaans Univ., Johannesburg, 1990).
- de Ronde, C. E. J. & de Wit, M. J. *Geology* (submitted).
- de Wit, M. J., Hart, R. A. & Hart, R. J. *J. Afr. Earth Sci.* **6**, 681–730 (1987).
- de Wit, M. J. & Tredoux, M. *Geoplatinum* **87**, 319–342 (Elsevier, London, 1988).
- Tredoux, M. et al. *J. geophys. Res.* **94**, 795–813 (1989).
- de Wit, M. J. *J. Afr. Earth Sci.* **13**, 45–63 (1991).
- Armstrong, R. A., Compston, W., de Wit, M. J. & Williams, I. S. *Earth planet. Sci. Lett.* **101**, 90–106 (1990).
- Dia, A., Allègre, C. J. & Erlank, A. J. *Earth planet. Sci. Lett.* **98**, 74–89 (1990).
- Durrheim, R. J. & Mooney, W. D. *Geology* **19**, 606–609 (1991).
- Van Reenen, D. D., Roering, C., Ashwal, L. D. & de Wit, M. J. *Precamb. Res.* **55**, 1–5 (1992).
- Roering, C. et al. *Precamb. Res.* **55**, 539–552 (1992).
- Roering, C. et al. *Precamb. Res.* **55**, 69–80 (1992).
- Coward, M. P. & Fairhead, J. D. *Tectonophysics* **68**, 31–43 (1980).
- de Wit, M. J., Armstrong, R. A., Kamo, S. & Erlank, A. J. *Econ. Geol.* (submitted).
- Hollister, L. S. *The Limpopo Belt: A Field Workshop on Granulites and Deep Crustal Tectonics* (ext. abstr.) 66–68 (ed. Barton, J. M. (Rand Afrikaans Univ., Johannesburg, 1990).
- Reimer, T. O. *Precamb. Res.* **12**, 393–410 (1980).
- de Wit, M. J. & Hart, R. A. in *Geochemical Evolution of the Continental Crust* (ext. abstr.) 196–202 (1988).
- de Wit, M. J., Fripp, R. E. P. & Stanistreet, I. G. *Geol. Soc. S. Afr. spec. Publ.* **9**, 21–30 (1983).
- Compston, W. & Kröner, A. *Earth planet. Sci. Lett.* **87**, 13–28 (1988).
- Kröner, A. & Todt, W. *J. geophys. Res.* **93**, 15329–15337 (1988).
- Hunter, D. J. *J. Afr. Earth Sci.* **13**, 13–26 (1991).
- Hart, R. A., Pyle, D. & de Wit, M. J. *Geochemistry and Mineral Formation in the Earth Surface* (eds Rodriguez-Clemente, R. & Tardy, Y.) 77–79 CSIC (Paris, Imp. Aguirre, 1987).
- de Wit, M. J., Tredoux, M. & Hart, R. A. *EOS* **71**, 969–970 (1990).
- de Ronde, C. E. J. thesis, Univ. of Toronto (1990).
- de Ronde, C. E. J., de Wit, M. J. & Spooner, E. T. C. *Bull. Geol. Soc. Am.* (submitted).
- de Wit, M. J., Hart, R. A., Martin, A. & Abbott, P. *Econ. Geol.* **77**, 1783–1802 (1982).
- Lowe, D. R. A. *Rev. Earth planet. Sci.* **8**, 145–167 (1980).
- Byerly, G. R. & Palmer, M. R. *Contrib. Min. Petr.* **107**, 387–402 (1991).
- de Wit, M. J., Armstrong, R., Hart, R. J. & Wilson, A. M. *Tectonics* **6**, 529–549 (1987).
- de Ronde, C. E. J., Kamo, S., Davis, D. W., de Wit, M. J. & Spooner, E. T. C. *Precamb. Res.* **49**, 261–280 (1991).
- Kamo, S., Davis, D. W. & de Wit, M. J. *Int. Conf. Geochron. Cosmochron. Isotope Geol.* (abstr.) *Geol. Soc. Austr.* **27**, 53 (1990).
- Anhaeusser, C. R. & Robb, L. J. *Geol. Soc. S. Afr. Spec. Publ.* **9**, 131–152 (1983).
- Toulkeridis, T., Goldstein, S. L., Kröner, A., Lowe, D. R. & Schidlowski, M. *3rd Int. Archaean Symp.* (ext. abstr.) (eds Glover, J. E. & Ho, S. E.) 309–310 (Geoconferences, 1990).
- Allsopp, H. L., Viljoen, M. J. & Viljoen, R. P. *Geol. Rundschau* **62**, 902–917 (1973).
- de Ronde, C. E. J., Hall, C. M., York, D. & Spooner, E. T. C. *Geochim. cosmochim. Acta* **55**, 1933–1951 (1991).
- Hunter, D. R. & Wilson, A. M. *S. Afr. J. Geol.* **91**, 57–74 (1988).
- Wilson, A. H. & Carlson, R. W. *Earth planet. Sci. Lett.* **96**, 89–105 (1989).
- de Wit, M. J., Jones, M. & Buchanan, D. L. *Precamb. Res.* **55**, 123–153 (1992).
- Groenewald, P. B., Grantham, G. H. & Watkeys, M. K. *J. geol. Soc. Lond.* **148**, 1115–1123 (1991).
- Richardson, S. H., Gurney, J. J., Erlank, A. J. & Harris, J. W. *Nature* **30**, 198–200 (1984).
- de Wit, M. J. *Lunar planet. Inst. Houston, Tech. Rep.* 86–88, 14–17 (1986).
- Rapp, R. P., Watson, E. B. & Miller, C. F. *Precamb. Res.* **51**, 1–25 (1991).
- Arndt, N. T. & Nisbet, E. G. (eds) *Komatiites* (Allen and Unwin, London, 1982).
- Basaltic Volcanism Study Project *Basaltic Volcanism on the Terrestrial Planets* (Pergamon, New York, 1981).
- Campbell, I. H. & Taylor, S. R. *Geophys. Res. Lett.* **10**, 1061–1064 (1983).
- Hamilton, P. J., O'Nions, R. K., Bridgwater, D. & Nutman, A. *Earth planet. Sci. Lett.* **62**, 263–272 (1983).
- Huang, X., Ziwei, B. & De Paolo, D. J. *Geochim. cosmochim. Acta* **50**, 625–631 (1986).
- Hofmann, A. W., Jochum, K. P., Seufert, M. & White, W. M. *Earth planet. Sci. Lett.* **79**, 33–45 (1986).
- Hart, R. J., Andreoli, M. A. G., Tredoux, M. & de Wit, M. J. *Chem. Geol.* **82**, 21–50 (1990).
- Burke, K. C., Kidd, W. S. F. & Kusky, T. M. *Tectonics* **5**, 439–456 (1986).
- de la Winter, H. S. *Afr. J. Geol.* **90**, 409–427 (1987).
- Roering, C., Barton, J. M. & de la Winter, H. *Tectonophysics* **171**, 7–22 (1990).
- Burke, K., Kidd, W. S. F. & Kusky, T. M. *J. Geodynam.* **2**, 35–49 (1985).
- Armstrong, R. A., Compston, W., Retief, E. A., Williams, I. S. & Welke, H. J. *Precamb. Res.* **53**, 243–266 (1991).
- Robb, L. J., Davis, D. W. & Kamo, S. L. *S. Afr. J. Geol.* **94**, 86–95 (1991).
- de Beer, J. M. & Eglinton, B. M. *J. Afr. Earth Sci.* **13**, 27–44 (1991).
- Matthews, P. E. *Crustal Evolution and Orogeny* (ed. Sychanthavong, S. P. M.) 41–73 (Oxford and IBH, New Delhi, 1990).
- Dietz, R. S. *J. Geol.* **61**, 499–516 (1961).
- Grieve, R. A. F., Coderre, J. M., Robertson, P. B. & Alexopoulos, J. *Tectonophysics* **717**, 185–200 (1990).
- Hart, R. J., Andreoli, M. A. G., Reimold, W. U. & Tredoux, M. *Tectonophysics* **192**, 313–331 (1991).
- Durrheim, R. J. *Am. geophys. Un. Geodyn. Ser.* **13**, 77–83 (1986).
- Hart, R. J., Andreoli, M. A. G., Smith, C. B., Otter, M. L. & Durrheim, R. *Chem. Geol.* **83**, 233–248 (1990).
- Roering, C. *Trans. Geol. Soc. S. Afr.* **87**, 87–89 (1984).
- Roering, C. & Smith, C. A. *J. struct. Geol.* **9**, 419–427 (1987).
- de Wit, M. J., van Reenen, D. D. & Roering, C. *Precamb. Res.* **55**, 111–122 (1992).
- McCourt, S. & Vearncombe, J. R. *Precamb. Res.* **55**, 553–570 (1992).
- Sengor, C. *Earth Sci. Rev.* **28**, 1–200 (1990).
- Stuart, G. W. & Zengeni, T. G. *Tectonophysics* **144**, 323–335 (1987).
- Babel Working Group *Nature* **348**, 34–38 (1990).
- Vearncombe, J. *Geol. Mag.* **123**, 601–609 (1986).
- Corner, B., Durrheim, R. J. & Nicolaysen, L. O. *Tectonophysics* **171**, 49–61 (1990).
- Wilson, J. F. et al. *Nature* **271**, 23–27 (1978).
- Vearncombe, J. *Tectonics* **7**, 761–774 (1988).
- Ericksson, K., Kidd, W. S. F., Krapex, B. in *New Perspectives in Basin Analysis* (Springer, New York, 1988).
- Minter, W. E. L. & Loen, J. S. *S. Afr. J. Geol.* **94**, 70–85 (1991).
- Stanistreet, I. G. & McCarthy, T. S. *J. Afr. Earth Sci.* **13**, 65–82 (1991).
- Howell, D. G. *Tectonics of Suspect Terranes* (Chapman and Hall, London, 1989).
- Oliver, J. *Geology* **14**, 99–102 (1986).
- Duane, M. J. & de Wit, M. J. *Geology* **16**, 999–1002 (1988).
- Dewey, J. F. *Tectonics* **7**, 1123–1139 (1988).
- Nur, A. & Ben-Avraham, Z. *The Evolution of the Pacific Ocean Margins* (ed. Ben-Avraham, Z.) 7–19 (Oxford Univ. Press, 1989).
- Abbott, D. H. & Hoffman, S. E. *Tectonics* **3**, 429–448 (1984).
- Nixon, P. H. & Neal, C. R. in *Mantle Xenoliths* (ed. Nixon, P. H.) 335–345 (Wiley, Chichester, 1987).
- Storey, M., Mahoney, J. J., Kroenke, L. W. & Saunders, A. D. *Geology* **19**, 376–379 (1991).
- Reymer, A. & Schubert, G. *Tectonics* **3**, 63–77 (1984).
- Hart, R., Welke, H. & Nicolaysen, L. O. *N. J. geophys. Res.* **86**, 10663–10680 (1981).
- Matthews, P. E. *Nature Phys. Sci.* **240**, 37–39 (1972).
- Scott, D. J., St-Onge, M. R., Lucas, S. B. & Helmstaedt, H. in *Ophiolite Genesis and Evolution of the Oceanic Lithosphere* (ed. Peters, T.) 825–857 (Kluwer, Dordrecht, 1991).
- Scott, D. J., Helmstaedt, M. & Bickle, M. J. *Geology* **20**, 173–176 (1992).

ACKNOWLEDGEMENTS. We thank many colleagues for discussions over the past 10 years, in particular L. Nicolaysen. The work was funded by the FRD/CSIR and NSF, and by the Anglo-American Corp Ltd. We thank P. Hoffman and A. Hynes for their critical comments, N. Adonis for typing and E. Vitali for producing Fig. 1a.

Gustducin is a taste-cell-specific G protein closely related to the transducins

Susan K. McLaughlin, Peter J. McKinnon & Robert F. Margolske*

Roche Institute of Molecular Biology, Roche Research Center, Nutley, New Jersey 07110-1199, USA

A novel G protein α -subunit (α -gustducin) has been identified and cloned from taste tissue. α -Gustducin messenger RNA is expressed in taste buds of all taste papillae (circumvallate, foliate and fungiform); it is not expressed in non-sensory portions of the tongue, nor is it expressed in the other tissues examined. α -Gustducin most closely resembles the transducins (the rod and cone photoreceptor G proteins), suggesting that gustducin's role in taste transduction is analogous to that of transducin in light transduction.

VERTEBRATE taste transduction is mediated by specialized taste receptor cells typically organized into groups of 40–60 to form the taste buds^{1–4}. Collections of lingual taste buds are present in three different types of specialized structures: the circumvallate, foliate and fungiform papillae. Taste transduction is initiated at the taste pore where the microvilli of the taste receptor cells make contact with the outside environment. Various taste ligands either depolarize or hyperpolarize taste cells to regulate neurotransmitter release from taste cells. Lateral connections between taste cells in a bud may also modulate the output transmitted to the afferent taste fibres.

There are four different taste submodalities, typified by four distinct groups of taste stimuli: salty, sour, sweet and bitter. But the molecular mechanisms underlying taste transduction have only been partially elucidated (for recent reviews see refs 5–7). Each taste submodality has its own specific mechanism: salty taste seems to be due to Na^+ flux through apical Na^+ channels^{8,9}; sour seems to be mediated by H^+ ion blockade of K^+ or Na^+ channels^{10,11}; bitter and sweet seem to be mediated by guanine nucleotide binding protein (G protein)-dependent mechanisms^{5–7}. Sweet compounds cause a GTP-dependent generation of cyclic AMP in rat tongue membranes^{12,13}. External application or microinjection of cAMP inactivates K^+ channels in vertebrate taste cells and leads to their depolarization^{14–16}. Furthermore, very high levels of adenylyl cyclase and cAMP phosphodiesterase are found in taste tissue^{17–19}. These findings suggest a transduction pathway in which the activation of a sweet receptor leads to cell depolarization through a gustatory G protein-mediated rise in cAMP. Bitter, at least in part, is also transduced through a G-coupled receptor: the bitter compound denatonium leads to Ca^{2+} release from internal stores²⁰ (probably by way of gustatory G protein-mediated inositol trisphosphate generation).

G proteins are heterotrimeric GTP-binding proteins that couple receptors to effectors in many diverse signal transduction processes (for recent reviews, see refs 21–24). The α -subunit of the G protein confers most of the specificity of interaction between the receptor and the effectors in the signal transduction process. There is a wide variety of G protein α -subunits: some are ubiquitously expressed (for example α_s and α_i), whereas other G proteins involved in sensory transduction are only expressed in specialized sensory cells. For example, the transducins (G_t) transduce photoexcitation in retinal rod and cone cells^{25–29}, and G_{olf} transduces olfactory stimulation in neurons of the olfactory epithelium³⁰.

We set out to identify proteins involved in vertebrate taste transduction: our initial focus was on G protein α -subunits. The experimental evidence cited above implicates G proteins

in the transduction of sweet and bitter. Furthermore, G protein α -subunits are highly conserved, making it possible to design specific primers that can be used in the polymerase chain reaction (PCR) to amplify and then clone their DNAs³¹. The identification of gustatory-specific G proteins will provide insight into the mechanisms of the taste pathways. We report here the cloning and identification of α -gustducin, a novel G protein α -subunit specifically expressed in taste cells.

Sequence of α -gustducin

To identify G proteins present in taste tissue, degenerate oligonucleotide primers from conserved regions of the G protein α -subunits were used in the PCR with taste tissue complementary DNA as the template. In this way, six different types of α -subunit clones were isolated from the taste cDNA library (Table 1): five of the α -subunit clones (α_{i2} , α_{i3} , α_{i4} , α_{i2} , α_s) have been previously identified and are expressed in several tissues other than lingual epithelium^{21–24,31–34}. In addition to these known α -subunits, we isolated a novel clone: we demonstrate below that this new α -subunit is taste-cell specific.

The complete sequence of the gustatory α -subunit (Fig. 1a) contains 1,703 base pairs (bp) of DNA with a single long open reading frame sufficient to encode a protein of 354 amino acids. Comparing the nucleic-acid sequence and the deduced amino-acid sequence of this gustatory α -subunit to that of other α -subunits demonstrates that it is a member of the α_i type superfamily (α_i , α_t , α_o , α_z).

This gustatory α -subunit is most closely related to the transducins (Table 2). Because of its close relationship to the transducins and its presumptive role in taste transduction we have named this gustatory G protein gustducin. At the amino-acid level the α -subunit of rat gustducin is 79–80% identical and 90% similar to the bovine α -transducins. For comparison, bovine rod α -transducin is 81% identical and 90% similar to bovine cone α -transducin. Because the rat α -transducin DNA sequences have not been determined, we cannot compare rat α -gustducin to the rat α -transducins. But among mammals, 1 to 3% differences in amino-acid identity are typical among α -isotypes, suggesting that the α -subunits of gustducin and the transducins comprise a subfamily of closely related proteins. In contrast, gustducin is only 66–68% identical to the α_i subunits, and only 46% identical to α_s (a similar level of relatedness exists between the transducins and α_i or α_s).

An alignment of α -gustducin with the α -subunits of bovine rod and cone transducin shows that the general structure of all three α -subunits is highly conserved (Fig. 1b). In fact, the carboxy-terminal 38 amino acids of α -gustducin and both transducins are identical. This C-terminal identity is of particular importance, because this region has been implicated in G protein/receptor interactions^{35,36}. Furthermore, the region from

* To whom correspondence should be addressed.

a

```

1  gactgggtgcctgctgttgggagcactgctctgacgatctatctctaaccactgctgtgctctgtgtttgaaaactttgagcaaatcaactgcccgctci
101 ctaacagcaaaagATGGGAAGTGAATTAAGTTCAGAGAGCAAGGACTCAGCCAAAAGGTCCAAAGAACTGGAGAAGAAGCTTCAGGAAGATGCTGAACGA
    M G S G I S S E S K E S A K R S K E L E K R L Q E D A E R
201 GATGCAAGAAGTGTGAAGTGTCTGCTATTAGGAGCAGGTGAATCTGGAAAGTACTATTTGTAAACAAATGAAGATCTCCACAGAAGTGGTTACACTA
    D A R T V K L L L G A G E S T G K S T I V R Q M K I I H K R N G Y C S R
301 AACAGAATGATGAGTGTAAAGCAGTGTCTTACAGTAACACGTGTCAGTCCATCTGSCCATTTGTGAAGCCATGACTACACTAGGGATTGATTATGT
    Q E C H E F T K A V V Y S N T L Q S I L A I V K A M T T L A G I D Y V
401 CAATCCGAGAAAGTAGAGAGGACCAACCTGCTTCTCTCCATGGCAAAAGCACTAGAAAGATGCGTGACATGACGCTCAGTGGCTGAAATTAATTAACGT
    N F R R S R E D Q A Q L L L S M A N T L C T A G A A G A T G D M T A C G C T C A G T G G C T G A A A T A A T I K R G T
501 CTGTGGGCGCATCCAGGAATTCAGCCTGCTTCGAAAGGCGCATCTGAATACCAGCTCAATGACTCTGCACTTACTACCTTAATGACTTATGATAGACTCA
    L W G G C D F C G A G A I T Q A C F T E R A S E Y Q L N B S A A Y A Y L N D L D A G A C T A
601 CAGCCCTCGGTGTGCTGCAAGAACAGCTTCTACATTCGCGGTGAAACCACTGATGATGAACTCAATCTCTCTTAAAGACTTGAACCTT
    A F G V Y F N E Q L L H S M A N T L C T A G A A G A T G D M T A C G C T C A G T G G C T G A A A T A A T I K R G T
701 CAGAAATGTTGATGTAGTGTGCGCAGAGATCAGAAAGAAAGAAATGGATCCTACTGCTTTGAGAGAGTGCATGATATATTTTGTGAGCAGCCCTAAGTGC
    R M F T D V G G Q R S E R R K A K W I H C F T E G G V T C I I F T C A A L S A
801 TACGACATGCTACTTGTGAAGATGAAGAGGTGAACAGAAATGCTTCAACAGCATCTGTAATCAAGTATTTTGAACCAACCTT
    Y D E V L V E D E E V N R M H E S L F H L N S I C N H R A T T T G C A A C C A C C T
901 CCATTGTTCTGTTTCTTAACAGAAAGATCTCTCCAGGAGAAAGTGACCAAGGTGCACCTCAGCATCTGTTTCCCAAGTACACTGGAACCAATACAT
    I V L F L T N K K D L F T Q E K V T K V H L S I C F T F E Y T G F N T F
1001 CGAAGATGCAGGAAGTACATCAAGAACCGTTCCTAGACCTGAACCTAAAGAAAGAGATAAGGAAATCTATTTCACATGACCTCGCTACTGACAC
    E D A C S N A Y I K N A C G T T C T A G A C T L A A K K E D K E I Y S H M C I A T D T
1101 CAAAACGTCAAATTCGTGTTTGTATGCCGTGACAGATATAATAATAAGAAAGAACCTCAAGAGCTGTGGGCTCTTCTGAGcaacctgttctaccactgt
    Q N V K A F T V F D A V T D I I I R E N L K A B C G L F T C T G A g c a a c c t g t t g c t a c c a c t t g
1201 tgatggctaagtctcttttaagacataaaaagggtgctgtgtattagctgtgtagatatiaactgatttagaaatgtgaactagcattataaaacaaaaa
1301 attcacacaaaatattactgtgatatacgtatatactgggtacgggtttcttggggaatggagggtagagtgtgtagtcttaaatctgaaatctgat
1401 gtatctggttaactgtcacaatatatacattcatgctactaaagtttttgggaagtgcgtgtgaagtaccaaatttttaatacatagagtaaaactcagaatgt
1501 gctataacaattgccccagcfagattttgaagacattcaaagtcatgtctgtactacagaaactgtacaaaatgaacaaggatataatttgggtcatcagcc
1601 tttcaattagctgccacaagcacacacagtagatgcttttattgatgggaaattgtatgtgtaaaataaatatataataaaaaa
1701 aaa 1703

```

FIG. 1 a, Nucleotide and deduced amino-acid sequence (single-letter code) of rat α -gustducin, a taste-cell-specific G protein α -subunit. Nucleotides are numbered in the 5' to 3' direction beginning with the first nucleotide of the cDNA clone. The inferred start of translation begins at nucleotide 114, with the first in-frame stop codon at nucleotide 1,176, yielding an open reading frame of 1,064 nucleotides, encoding a 354-amino-acid polypeptide. The nucleic-acid sequence and inferred amino-acid sequence of α -gustducin demonstrate that it is a G protein α -subunit in the α_i superfamily (α_i , α_1 , α_o and α_2). Potential sites are present for pertussis toxin (C351) and cholera toxin (R178) mediated ribosylation. A potential *N*-myristoylation site (MGXXXS) is present at the N terminus of α -gustducin. The complete DNA sequence was determined for two independent clonal isolates (T93 and T95). Clone T93 was of partial length, lacking 5' end sequences and was apparently unspliced containing an insertion of 193 bp between nucleotides 833 and 834. Clone T95 was deleted for sequences 704 to 838, because of apparent spliceout of an exon: the DNA sequence of this region was derived from clone T93 and the products of several different PCR amplifications of library cDNA using primers spanning this region (for example, TIVKQM to VFDAVTD). The DNA sequence displayed is a composite of clone T95, T93 and PCR products from TIVKQM to VFDAVTD. **b**, Alignment of amino-acid sequences of α -subunits of rat gustducin, bovine cone transducin (bovine cone) and bovine rod transducin (bovine rod). Amino-acid sequence alignments were produced iteratively by the BestFit routine of the Wisconsin GCG software package⁵⁶. Consensus sequence matches (that is, at least two out of the three proteins match) are denoted by white letters in black boxes. Conservative changes are denoted by black letters in shaded boxes. Non-conservative changes are depicted by black letters in unboxed regions. The consensus line shows positions where two or three of the three proteins match; dashes in the consensus sequence correspond to

non-conserved positions. Dots in the rod transducin sequence correspond to gaps to align it with the other sequences. Note the high degree of conservation of the three sequences throughout their length. The last 38 amino acids of all three proteins are identical: this region has been implicated in receptor interaction. A large number of differences are clustered in the region from V96 to S109 of gustducin: this is a highly variable region of G protein α -subunits. The high degree of conservation among the three proteins (~80% identity and ~90% similarity) suggests that their enzymatic properties (GTP binding and GTPase activity) and receptor-effector interactions are likely to be quite similar.

METHODS. Primers described in the legend to Table 1 (KWIHCF, FLNKKD, HLFNSIC, VFDAVTD and TIVKQM) and other oligonucleotides derived from conserved regions of the G protein α -subunits were used in the PCR (under conditions described in Table 1 legend). α -Gustducin-specific clones were recovered from PCR amplifications with KWIHCF and FLNKKD; HLFNSIC and VFDAVTD; TIVKQM and FLNKKD; TIVKQM and KYFATTS; T7 sequencing primer (Stratagene) and LAEIKR. The DNAs were sequenced as described in Table 1 legend. The following non-degenerate PCR primers corresponding to α -gustducin sequences were made: **KYFATTS** [CCGGATCCGAGGTGG-TTGCAAAATACTT] (882); **LAEIKR** [CGGATCCGACGTTAATTATTTAG-CCAA] (480). The underlined sequences at the 5' end of each oligonucleotide contain a restriction endonuclease site to aid cloning. The number in parentheses at the end of each sequence refers to the α -gustducin nucleotide location for the codon corresponding to the first amino acid of the primer (see **a**). An α -gustducin clone isolated by PCR amplification with TIVKQM and KYFATTS was used to screen the taste cell enriched cDNA library for α -gustducin specific clones. Four such clones (T95, T93, T85 and T77) were identified and their DNAs sequenced. Clones T93, T85 and T77 were similar in structure, clone T95 was distinct.

Q137 to F354 is extremely similar for all three subunits: in this region each α -subunit differs from the consensus sequence at only 14 or 15 positions. This region also contains most of the sites implicated in guanine nucleotide binding^{37,38}. Specific amino acids that are known to be implicated in GTPase activity include G42, R197 and Q204: all of these amino acids are conserved in these three proteins. These comparisons indicate that the guanine-nucleotide-binding properties and GTPase activities of these three α -subunits are likely to be quite similar. All three α -subunits contain a potential *N*-myristoylation site

(MGXXXS) at their amino terminus, which if used may anchor these α -subunits to the inner face of the plasma membrane.

Transcription of the gustducin gene

Although α -gustducin is closely related to the transducins, it differs from them at the amino-acid level at several positions scattered throughout its sequence. This suggests that α -gustducin is transcribed from a gene distinct from the transducins. Southern blot analysis (Fig. 2a) with α -gustducin probes and α -transducin probes confirms this: α -gustducin is a single-copy

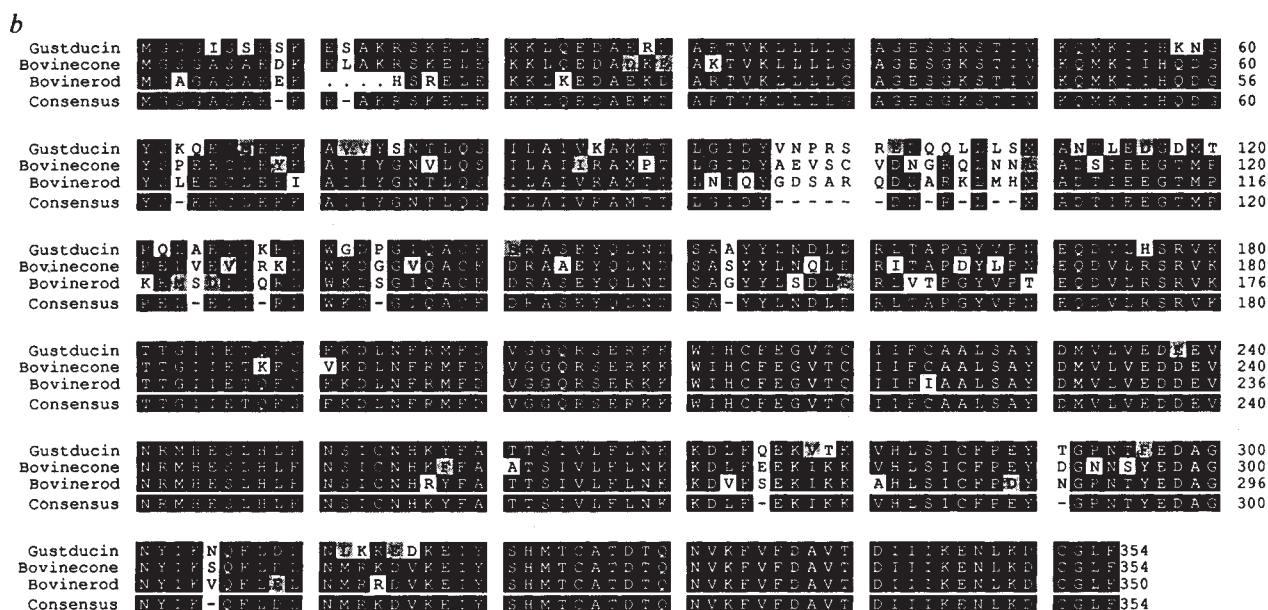
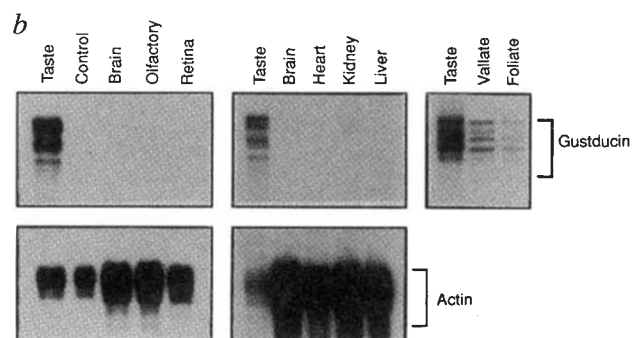


FIG. 2 a, Southern blot analysis of rat genomic DNA with α -gustducin and α -transducin probes. DNAs were digested with the various restriction endonucleases indicated at the top of each lane. **b**, Left panels, tissue-specific expression of α -gustducin transcripts as assayed by RNase protection of *in vitro* generated RNA. The cDNA libraries from various tissues were prepared, linearized with restriction endonuclease, then *in vitro* run-off transcripts were generated from the total cDNA libraries. RNase protection was done simultaneously with α -gustducin (upper panels) and actin probes (lower panels) to normalize for expression. α -Gustducin transcripts are only present in RNA from the taste cell enriched library (taste), and absent from libraries derived from non-taste lingual epithelium (control), retina, brain and olfactory epithelium. Middle panels, RNase protection with total RNA (15 μ g) from various tissues demonstrates that α -gustducin is absent from brain, heart, liver and kidney; total RNA from the pooled circumvallate and foliate papillae (taste) of three rats demonstrates the presence of α -gustducin in this tissue. Right panel, RNase protection with *in vitro* generated RNA (2 μ g) from the taste cell enriched library (taste) against total RNA derived from nine circumvallate papillae (vallate) or nine foliate papillae demonstrates that α -gustducin is present in both foliate and circumvallate papillae derived total RNA.

METHODS. For Southern blots, genomic DNA was isolated from rat liver⁵⁷. DNA (40 μ g) was digested to completion with 20 units of restriction endonuclease, electrophoresed in a 1.0% agarose gel, then transferred to nitrocellulose. The nitrocellulose membrane was cut in half (longitudinally), then probed with gustducin- or transducin-specific probes. Specific DNA fragments of gustducin and transducin (rat rod type) were cloned after PCR amplification from rat cDNA using the TIVKQM and KYFATS primers (gustducin), and the TIVKQM and LVEDDEV [CCGGATCCCTTCGTCGCTCCACCAGC] primers (transducin). The most divergent regions (bp 255–526 of gustducin and the corresponding region of rod transducin) were subcloned from these clones and used as probes. Radioactive labelled probes were generated by PCR amplification during which all of the dCTP present was α -³²P-labelled. Unincorporated nucleotides

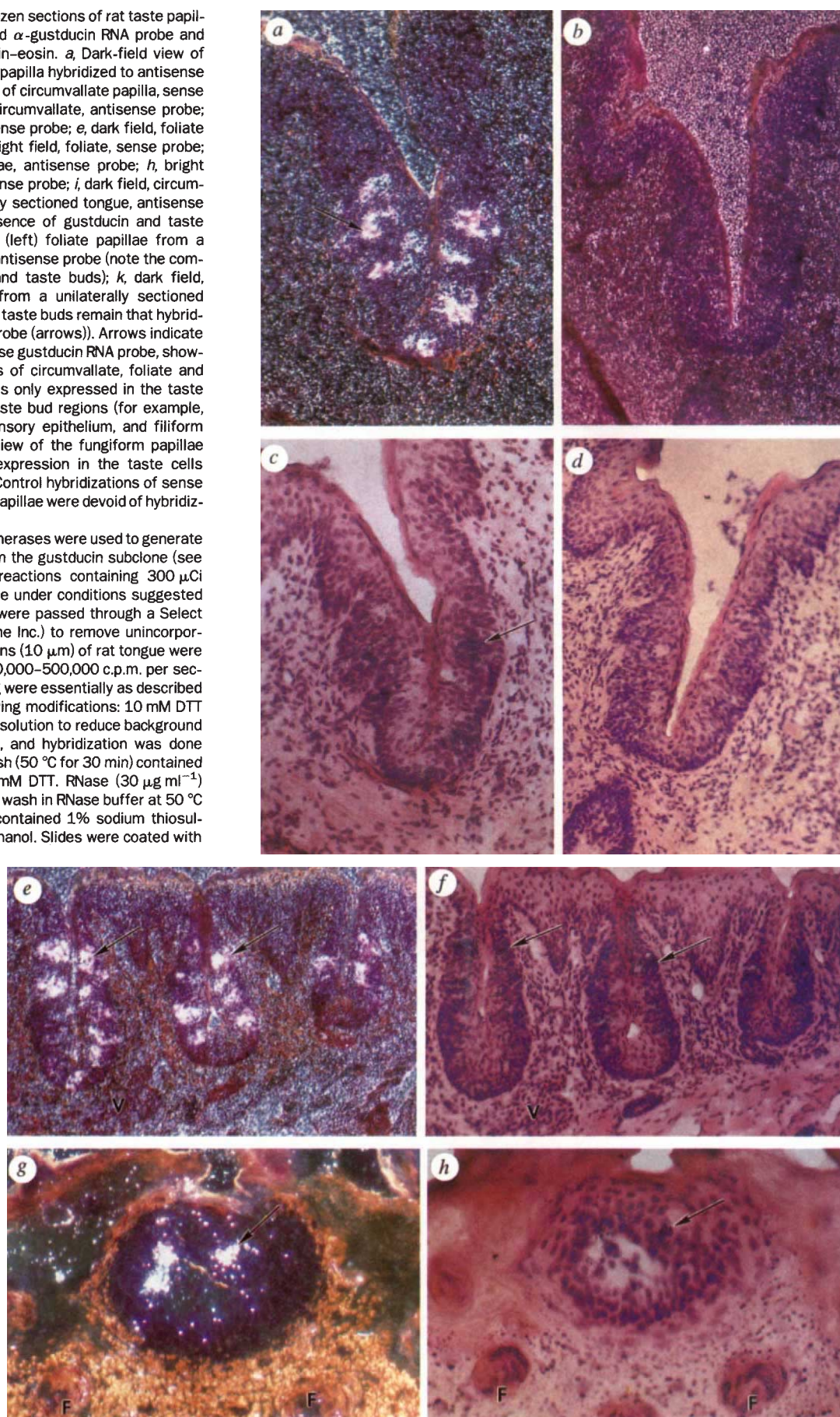


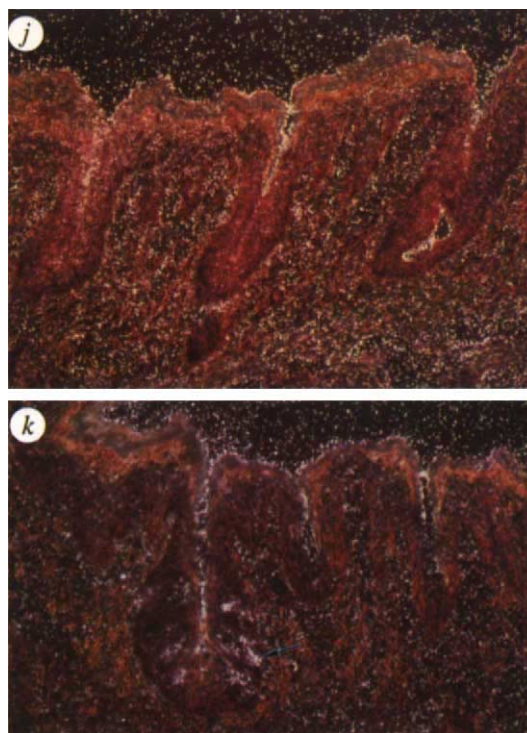
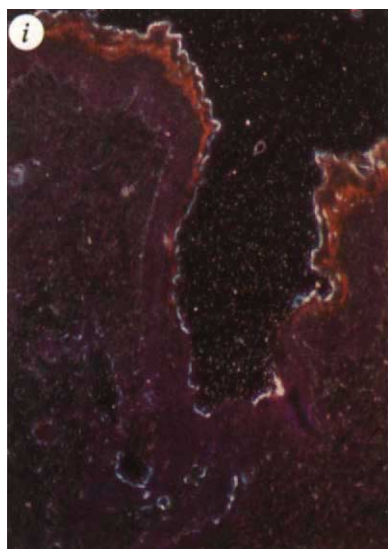
were removed from the probe using a G-50 spin column (5 Prime $\rightarrow$ 3 Prime Inc.). The blots were hybridized with 5×10^7 c.p.m. of probe in 20 ml 50% formamide hybridization solution at 42 $^{\circ}$ C overnight. Hybridization and washing conditions were as detailed⁵⁸. For RNase protection assays, poly(A)⁺ RNA was prepared as described in the legend to Table 1. Total RNA was collected from taste papillae using 4 M guanidinium thiocyanate lysis followed by organic extraction and differential precipitation as described⁵⁷. *In vitro* sense RNAs were generated by T7 RNA polymerase (USB) mediated transcription of library DNAs which had been linearized with endonuclease *Not*I. Sense RNA (75–100 μ g) was generated from 25 μ g of linearized library DNA. ³²P-labelled antisense RNAs were transcribed by T3 RNA polymerase (Boehringer). The template for gustducin transcripts was a clone derived by the PCR with primers KWIHCF and FLNKKD (nt 741 to 929). The actin template contained a portion of the rat cytoplasmic β -actin gene (GenBank accession number J00691) (nt 2,469–2,588). The β -actin clone was generated by PCR of the non-taste lingual cDNA library using the following β -actin-specific primers: CCGATCCGAGATGGCCACTGCCGCAT and AGAATTCGGAAGGCTGGAAGAGAGC. The product of this PCR reaction was then cloned as a 120 bp *Eco*RI–*Bam*HI fragment into pBS KSII⁺ (Stratagene). Antisense actin RNA was generated from *Bam*HI endonuclease linearized pBS-actin DNA using T3 RNA polymerase. All RNase protection assays were done using a kit from Ambion according to the method of Krieg and Melton⁵⁹. About 500 c.p.s. each of ³²P-labelled antisense gustducin and actin RNAs were combined with 2 μ g antisense RNA or 15 μ g total RNA. The total RNA from taste papillae was the total yield from three animals and was not quantified. The total RNA for the circumvallate and foliate papillae corresponded to nine papillae each and was not quantified. Protected products were analysed on a 5% acrylamide–urea sequencing gel and visualized by exposure on Kodak X-omat X-ray film at –80 $^{\circ}$ C using an intensifying screen.

FIG. 3 Photomicrographs of frozen sections of rat taste papillae hybridized with ^{35}S -labelled α -gustducin RNA probe and then stained with haematoxylin-eosin. *a*, Dark-field view of cross-section of circumvallate papilla hybridized to antisense α -gustducin probe; *b*, dark field of circumvallate papilla, sense probe control; *c*, bright field, circumvallate, antisense probe; *d*, bright field, circumvallate, sense probe; *e*, dark field, foliate papillae, antisense probe; *f*, bright field, foliate, sense probe; *g*, dark field, fungiform papillae, antisense probe; *h*, bright field, fungiform papillae, antisense probe; *i*, dark field, circumvallate papilla from a bilaterally sectioned tongue, antisense probe (note the complete absence of gustducin and taste buds); *j*, dark field, ipsilateral (left) foliate papillae from a unilaterally sectioned tongue, antisense probe (note the complete absence of gustducin and taste buds); *k*, dark field, contralateral foliate papillae from a unilaterally sectioned tongue, antisense probe (some taste buds remain that hybridize to α -gustducin antisense probe (arrows)). Arrows indicate area of hybridization to antisense gustducin RNA probe, showing its presence in taste buds of circumvallate, foliate and fungiform papillae. Gustducin is only expressed in the taste buds; it is absent from non-taste bud regions (for example, von Ebner's gland (V); non-sensory epithelium, and filiform papillae (F)). The bright-field view of the fungiform papillae (*g*) demonstrates gustducin expression in the taste cells extending into the taste pore. Control hybridizations of sense probe to fungiform and foliate papillae were devoid of hybridization (data not shown).

METHODS. T3 and T7 RNA polymerases were used to generate ^{35}S -labelled RNA probes⁵⁹ from the gustducin subclone (see Fig. 2 legend). Transcription reactions containing 300 μCi ^{35}S -labelled thio-UTP were done under conditions suggested by Promega. The RNA probes were passed through a Select DR-F column (5 Prime $\rightarrow$ 3 Prime Inc.) to remove unincorporated nucleotides. Frozen sections (10 μm) of rat tongue were hybridized with RNA probe (300,000–500,000 c.p.m. per section). Hybridization and washing were essentially as described by Lugo *et al.*⁶⁰ with the following modifications: 10 mM DTT was added to the hybridization solution to reduce background from ^{35}S -labelled RNA probes, and hybridization was done overnight at 50 $^{\circ}\text{C}$. The first wash (50 $^{\circ}\text{C}$ for 30 min) contained 50% formamide, 1 $\times$ SSC, 10 mM DTT. RNase (30 $\mu\text{g ml}^{-1}$) treatment was followed by one wash in RNase buffer at 50 $^{\circ}\text{C}$ for 30 min. All wash buffers contained 1% sodium thiosulphate and 0.1% β -mercaptoethanol. Slides were coated with

Kodak NTB-2 nuclear track emulsion as detailed⁶⁰, and were exposed at 4 $^{\circ}\text{C}$ for 3–4 weeks at which time they were developed, fixed and stained. Two types of denervation were done: (1) bilateral sections of both glossopharyngeal nerves and (2) unilateral section of the left glossopharyngeal nerve and the left chorda tympani. Fourteen days post-surgery (to allow full degeneration of taste buds) tissue sections containing foliate and circumvallate papillae were subjected to *in situ* hybridization with α -gustducin antisense probe, then stained with haematoxylin-eosin.





gene with a distinct restriction endonuclease digestion pattern.

In screening the taste tissue cDNA library, we found an apparent splicing variant of α -gustducin. One clone (T95) contained the entire coding region of α -gustducin, but had an in-frame deletion of 135 bp. When this cDNA sequence is aligned with the genomic sequence of murine α -transducin³⁹, the deletion corresponds to the precise removal of the sixth exon. The general exon-intron organization of the G protein α -subunits is highly conserved; therefore, it is likely that the deletion of clone T95 corresponds to splicing out of gustducin exon 6. Another clone from the cDNA library contained an insertion of 193 bp: comparison with the sequence of the exon-intron boundaries of murine transducin indicates that this insertion is due to the presence of an unspliced intron between exons 6 and 7. PCR reactions using primers spanning these exons show that the aberrantly spliced (deleted) variant and the unspliced form are present in taste cell cDNA at levels roughly one tenth that of the correctly spliced α -gustducin cDNA. The amino acids present in exon 6 (Arg 197 to Asn 241) are highly conserved for α_i and α -transducin and have been implicated in guanine nucleotide binding. If this deleted form of gustducin is actually produced it would differ significantly from other α -proteins in its guanine-nucleotide-binding properties and GTPase activities.

Expression of α -gustducin

Expression of α -gustducin mRNA was assayed by northern blot, primer extension and RNase protection. The RNase protection assays (Fig. 2b) demonstrate the presence of α -gustducin RNA only in taste tissue enriched preparations (poly(A)⁺ mRNA, total RNA or *in vitro* generated cDNA library sense RNA). No α -gustducin RNA was detected in the other tissues examined: non-taste lingual tissue, olfactory epithelium, retina, brain, liver, heart or kidney. RNase protection demonstrated the presence of α -gustducin messenger in both foliate- and circumvallate-derived total RNAs (Fig. 2b, right panel). Northern blot analysis of poly(A)⁺ mRNA from taste tissue indicates

three transcripts: a closely spaced doublet $\sim$ 1,700–1,800 nucleotides (nt) and a faint third band $\sim$ 1,500 nt (data not shown). Primer extension products had the same 5' terminus as clone T95 (data not shown). These results indicate that the full-length α -gustducin clone is $\sim$ 1,700 nt (Fig. 1b).

The RNase protection data demonstrate that expression of gustducin mRNA is restricted to tissue containing taste buds and absent from non-taste portions of the lingual epithelium. To determine directly if α -gustducin mRNA expression is confined to the taste buds, *in situ* hybridizations were done on tongue sections containing taste papillae. *In situ* hybridization demonstrates the presence of α -gustducin mRNA in the taste buds of circumvallate, foliate and fungiform papillae (Fig. 3a–h). Gustducin mRNA is completely absent from the adjacent lingual epithelium, muscle, connective tissue and von Ebner's glands. The sense probe controls show no hybridization to lingual tissue; the anti-sense gustducin probe shows no hybridization to retina (that is, cross hybridization between α -gustducin probe and retinal transducin mRNA did not occur; data not shown).

To determine if the expression of α -gustducin mRNA is dependent on the presence of taste buds, *in situ* hybridizations were carried out with frozen sections taken from rats whose tongues had been denervated. When the nerves to the taste buds are severed, the buds degenerate and do not reappear until after the connection is restored^{40,41}. If α -gustducin mRNA is present only in the taste buds, it follows that on degeneration of the taste buds, α -gustducin mRNA will disappear. The circumvallate papilla is innervated only by the glossopharyngeal nerves; bilateral sectioning of these nerves causes the taste buds of this papilla to degenerate. The foliate papillae are innervated by glossopharyngeal and chorda tympani nerves; unilateral sectioning causes the taste buds of the ipsilateral foliate papilla to degenerate, but leaves the taste buds of the contralateral foliate papilla intact.

After bilateral glossopharyngeal denervation, the circumvallate papilla was totally devoid of taste buds: α -gustducin mRNA

expression was also absent from the circumvallate papilla (Fig. 3i). As expected, taste buds expressing α -gustducin mRNA were still present in the foliate papillae of these rats (because input from the chorda tympani remained). But the number of taste buds in these papillae did seem to be reduced (data not shown). After unilateral sectioning of the left chorda tympani and left glossopharyngeal nerve, the ipsilateral foliate papilla was devoid of taste buds and displayed no detectable expression of α -gustducin mRNA (Fig. 3j). But the contralateral foliate papilla retained taste buds that did express α -gustducin (Fig. 3k). These results directly correlate the presence of innervated taste buds with α -gustducin mRNA expression.

Discussion

We have cloned a novel G protein subunit, α -gustducin, from rat taste tissue. This G protein subunit is only expressed in the

TABLE 1 Isolates of α -subunit clones from PCR-amplified rat taste tissue cDNA

	5'KWIHCF 3'FLNKKD	5'DVGGQR 3'FLNKKD	5'HLFNSIC 3'VFDAVTD	5'TIVKQM 3'FLNKQD
α_{i-2}	4	—	—	—
α_{i-3}	5	—	—	—
α_{14}	—	1	—	—
α_{12}	—	1	—	—
α_s	—	—	—	1
α_{gust}	5	—	4	1

The G protein α -subunit isolates cloned from PCR-amplified rat taste tissue cDNA are listed in the left-hand column. The assignment to α -subtype was made by comparing the DNA and inferred amino-acid sequences to data present in GenBank. The heading above each column lists the upstream (5') and downstream (3') primers. The numbers in each column represent independent clonal isolates from the particular PCR amplification. The circumvallate and foliate papillae from 90 Sprague-Dawley rats were collected⁵⁵ and immediately frozen in 100% ethanol at -70°C . An equivalent amount of non-taste lingual epithelium (devoid of taste buds) was also collected. We estimate that the taste buds comprised about 10–30% of the total mass of tissue collected. Poly(A)⁺ mRNA was isolated from taste and non-taste lingual tissue using the Quick Prep kit from Pharmacia: 7.9 μg mRNA was recovered from the taste tissue and 2.4 μg mRNA was recovered from the control non-taste lingual tissue. The BRL pSPORT vector and the BRL superscript kit were used to make two cDNA libraries from 1 μg of taste and 1 μg of non-taste lingual mRNA. The taste library contained 2.6×10^6 independent clones (average insert size of 1.1 kilobases (kb)). The non-taste library contained 4.8×10^6 independent clones (average insert size of 1.0 kb). Degenerate PCR primers corresponding to conserved amino acids of G proteins were made on an Applied Biosystems DNA synthesizer (single-letter amino-acid notation in bold typeface): **KWIHCF** [CGGATCCAA(AG)TGGAT(AC)TCA(CT)TG(CT)TT] (741); **FLNKKD** [(GGAATTC-(AG)T(CT)TT(CT)TT(AG)TT(AC)G)AG(AG)AA and GGAATTC(AG)T(CT)TT(CT)TT(AG)T(CT)AA(AG)AA] (912); **DVGGQR** [GTCTAGAGA(CT)GT(ACGT)GG-(ACGT)GG(ACGT)CA(AG)(AC)G] (711); **HLFNSIC** [CCGGATCCGCACCTGTTCAACAGCATCT] (855); **VFDAVTD** [CCGAATTC(ACGT)GT(ACGT)AC-(ACGT)GC(AG)TC(AG)AA(ACGT)AC] (1,116); **TIVKQM** [CCGAATTCAC(ACGT)AT-(ACGT)GT(ACGT)AA(AG)CA(AG)ATG] (255); **FLNKQD** [CCGAATTC(AG)TC(TC)-TG(TC)TT(AG)TT(ACGT)A(AG)(AG)AA] (912). The oligonucleotides corresponding to the 3' primers (FLNKKD, VFDAVTD, FLNKQD) were synthesized in the antisense orientation. The underlined sequences at the 5' end of each oligonucleotide contain a restriction endonuclease site to facilitate cloning. The number in parentheses at the end of each sequence refers to the first α -gustducin nucleotide location for the codon corresponding to the first amino acid of the primer (see Fig. 1a). Three of these primers (DVGGQR, KWIHCF and FLNKKD) have been previously used by Simon and co-workers³¹ in an earlier screen of G proteins. PCR samples contained 250 pmol of each primer and 20 ng of taste cell library cDNA in a 50 μl reaction volume. PCR followed this program: 94° for 1 min; 37° to 72° with a rise of time of 1° per 4 s; 72° for 3 min; for three cycles followed by: 94° for 1 min; 43° for 2 min; 72° for 3 min; for a total of 35 additional cycles. The PCR products were digested with *Bam*HI and *Eco*RI, electrophoresed in a 1% agarose gel, then bands of expected size were excised, purified and cloned into the pBluescript vector. After transformation into *Escherichia coli*, individual colonies were picked, and their DNAs sequenced using T7 DNA polymerase.

taste buds of the tongue: its mRNA is apparently present in all of the taste buds examined. Furthermore, α -gustducin mRNA is not expressed in two other sensory tissues (retina and olfactory epithelium) nor is it expressed in several non-sensory tissues (brain, liver, kidney, muscle and heart). The deduced amino-acid sequence of α -gustducin indicates that it is closely related to the α -transducins. We infer that they are evolutionarily related members of an α -subtype gene family. The close relationship of α -gustducin to the α -transducins implies that functional constraints have maintained this high degree of structural conservation.

We propose that gustducin's role in taste transduction is similar to that of transducin in visual transduction. In the retina the transducins relay activation of a receptor (rhodopsin or the colour opsins)^{42,43} into activation of a cytoplasmic effector enzyme (cGMP phosphodiesterase (PDE))^{44,45}. Transducin removes the inhibition from cGMP PDE by binding to the inhibitory γ -subunits; this leads to breakdown of cGMP and causes closure of cyclic nucleotide-gated channels^{46,47}. By analogy and because of structural conservation, we propose that α -gustducin transduces taste receptor activation into activation of a taste cell PDE. Previous studies have demonstrated extremely high levels of cAMP PDE in taste tissue^{17–19}. Perhaps α -gustducin binds to an inhibitory subunit of a taste cell cAMP PDE to activate this enzyme, and thereby decrease cAMP levels. We have cloned a catalytic subunit of a cAMP type PDE from our taste cell enriched library, and demonstrated that it is expressed in taste tissue, but is absent from non-sensory lingual tissue⁴⁸. At the level of the taste receptor we expect some functional and structural homology to the opsins. The C-terminal end of transducin is thought to interact with the opsins^{49,50}; the terminal 38 amino acids of α -gustducin and α -transducin are identical suggesting that gustducin can interact with opsin and opsin-like G-coupled receptors, and conversely that transducin can interact with taste receptors. The third cytoplasmic loop of rhodopsin is a key interaction site for transducin^{51,52}; we anticipate that the analogous region of the taste receptor will show structural and functional similarity.

Several biochemical^{12,13} and electrophysiological studies^{14–16} argue for sweet pathway involvement of G_s (or a G_s -like protein) which activates adenylyl cyclase to raise taste-cell cAMP levels. Transducin does not activate adenylyl cyclase, and gustducin by homology is unlikely to do so. Gustducin may inhibit the sweet response (PDE activation would decrease sweet-induced cAMP), or gustducin may play a role in bitter transduction. In support of these proposals, previous workers have correlated bitter compounds with PDE activation¹⁸, and several PDE

TABLE 2 Per cent sequence identity (amino acid/nucleic acid) among various G protein α -subunits

	Bov α_{t-rod}	Bov α_{t-cone}	Rat α_{gust}
Bov α_{t-rod}	—	81/73	80/71
Bov α_{t-cone}	81/73	—	79/73
Rat α_{gust}	80/71	79/73	—
Rat α_i	65/69	68/69	66/65
Rat α_o	62/66	62/65	62/63
Rat α_s	42/64	45/62	46/59

The TFasta and Fasta programs of the Wisconsin GCG software package⁵⁶ were used to search the GenBank for DNA and amino-acid sequences related to rat α -gustducin (the per cent identity at the amino-acid level is the left-hand number in each column; the right-hand number is per cent identity at the DNA sequence level). The two closest matches to α -gustducin are bovine rod (Bov α_{t-rod}) and bovine cone (Bov α_{t-cone}) transducins. A comparison of rat α -gustducin and the bovine transducins to other G protein α -subunits is presented in the three columns of the table. Note that the transducins are as closely related to each other (81% amino-acid identity) as they are to rat α -gustducin (79–80% amino-acid identity). These three G protein α -subunits are much less closely related to α_i , α_o and α_s .

inhibitors have been shown to enhance sweet perception^{53,54}. The biochemical and molecular biological characterization of taste transduction has previously been hindered because taste receptor cells comprise only a small inaccessible fraction of the

lingual epithelium. Using PCR we have cloned α -gustducin, a novel taste-cell-specific G-protein subunit. In like fashion, it should be possible to clone other components of the taste transduction pathways. □

Received 2 March; accepted 18 May 1992.

- Beidler, L.-M. in *Ciba Found Symp. Taste and Smell in Vertebrates* (eds Wolstenholme, G. E. W. & Knight, J.) 51–67 (Churchill, London, 1970).
- Graziadei, P. P. C. in *2nd Symp. Oral Sensation and Perception* (ed. Bosma, J. F.) 5–31 (Charles C. Thomas, Springfield, 1970).
- Reutter, K. Z. *Zellforsch.* **120**, 280–308 (1971).
- Grover-Johnson, N. & Farbman, A. *Cell Tissue Res.* **169**, 395–403 (1976).
- Kinnamon, S. C. *Trends Neurosci.* **11**, 491–496 (1988).
- Avenet, P. & Lindemann, B. *J. Membrane Biol.* **112**, 1–8 (1989).
- Roper, S. D. A. *Rev. Neurosci.* **12**, 329–353 (1989).
- Hack, G.-L., Mierson, S. & De Simone, J. A. *Science* **223**, 403–405 (1984).
- Schiffman, S. S., Lockhead, E. & Maes, F. W. *Proc. natn. Acad. Sci. U.S.A.* **80**, 6136–6140 (1983).
- Kinnamon, S. C. & Roper, S. D. *J. gen. physiol.* **91**, 351–371 (1988).
- Kinnamon, S. C., Dionne, V. E. & Beam, K. G. *Proc. natn. Acad. Sci. U.S.A.* **85**, 7023–7027 (1988).
- Striarn, B. J., Pace, U., Zehavi, U., Naim, M. & Lancet, D. *Biochem. J.* **260**, 121–126 (1989).
- Striarn, B. J. *et al. Chem. Senses* **15**, 529–536 (1990).
- Avenet, P. & Lindemann, B. *J. Membrane Biol.* **97**, 223–239 (1987).
- Tonosaki, K. & Funakoshi, M. *Nature* **331**, 354–356 (1988).
- Avenet, P., Hofmann, F. & Lindemann, B. *Nature* **331**, 351–354 (1988).
- Kurihara, K. & Koyama, N. *Biophys. Res. Commun.* **48**, 30–34 (1972).
- Price, S. *Nature* **241**, 54–55 (1973).
- Law, J. S. & Herkin, R. I. *Res. Commun. Chem. Pathol. Pharmacol.* **38**, 439–452 (1982).
- Akabas, M. H., Dodd, J. & Al-Awqati, Q. *Science* **242**, 1047–1050 (1988).
- Stryer, L. & Bourne, H. R. A. *Rev. Cell Biol.* **2**, 391–419 (1986).
- Gilman, A. G. A. *Rev. Biochem.* **56**, 615–649 (1987).
- Birnbaumer, L. A. *Rev. Pharmac. Tox.* **30**, 675–705 (1990).
- Simon, M. I., Strathmann, M. P. & Gautam, N. *Science* **252**, 802–808 (1991).
- Lochrie, M. A., Hurley, J. B. & Simon, M. I. *Science* **228**, 96–99 (1985).
- Medynski, D. C. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 4311–4315 (1985).
- Tanabe, T. *et al. Nature* **315**, 242–245 (1985).
- Yatsunami, K. & Khorana, H. G. *Proc. natn. Acad. Sci. U.S.A.* **82**, 4316–4320 (1985).
- Lerea, C. L., Somers, D. E., Hurley, J. B., Klock, I. B. & Bunt-Milam, A. H. *Science* **243**, 77–80 (1986).
- Jones, D. T. & Reed, R. R. *Science* **244**, 790–795 (1989).
- Strathmann, M., Wilkie, T. M. & Simon, M. I. *Proc. natn. Acad. Sci. U.S.A.* **86**, 7407–7409 (1989).
- Jones, D. T. & Reed, R. R. *J. Biol. Chem.* **262**, 14241–14249 (1987).
- Strathmann, M. & Simon, M. I. *Proc. natn. Acad. Sci. U.S.A.* **87**, 9113–9117 (1990).

- Pang, I.-H. & Sternweiss, P. C. *J. Biol. Chem.* **265**, 18707–18721 (1990).
- Navon, S. & Fung, B. *J. Biol. Chem.* **263**, 489–497 (1988).
- Hamm, H. *et al. Science* **241**, 832–835 (1988).
- Halliday, K. R. *J. Cyclic Nucl. Prot. Phosphorylation Res.* **9**, 435–448 (1984).
- Feig, A., Pan, B.-T., Roberts, T. M. & Cooper, G. M. *Proc. natn. Acad. Sci. U.S.A.* **83**, 4607–4611 (1986).
- Raport, C. J., Dere, B. & Hurley, J. B. *J. Biol. Chem.* **264**, 7122–7128 (1989).
- Whiteside, B. *J. comp. Neurol.* **40**, 33–45 (1976).
- Guth, L. *Expl. Neurol.* **8**, 336–349 (1963).
- Nathans, J. & Hogness, D. S. *Cell* **34**, 807–814 (1983).
- Nathans, J., Thomas, D. & Hogness, D. S. *Science* **232**, 193–202 (1986).
- Ovchinnikov, Y. A. *et al. FEBS Lett.* **204**, 288–292 (1986).
- Ovchinnikov, Y. A. *et al. FEBS Lett.* **223**, 169–173 (1987).
- Wensel, T.-G. & Stryer, L. *Biochemistry* **29**, 2155–2161 (1990).
- Fung, B. K. & Griswold-Prenner, I. *Biochemistry* **28**, 3133–3137 (1989).
- Spickofsky, N., McLaughlin, S. K., McKinnon, P. J. & Margolske, R. F. *Association for Chemoreception Sciences Abstracts* **14**, 192 (Sarasota, 1992).
- U, M. *et al. Adv. Cyclic Nucl. Prot. Phosphorylation Res.* **17**, 145–151 (1984).
- VanDop, C. *et al. J. Biol. Chem.* **259**, 23–26 (1984).
- Kuhn, H. in *Progress in Retinal Research* (eds Osborne, N. & Chader, J.) 123–152 (Pergamon, New York, 1984).
- Findlay, J. B. C. & Pappin, D. J. C. *Biochem. J.* **238**, 625–642 (1986).
- Schiffman, S. S., Gill, J. M. & Diaz, C. *Pharmac. Biochem. Behav.* **22**, 195–203 (1985).
- Schiffman, S. S., Diaz, C. & Beeker, T. G. *Pharmac. Biochem. Behav.* **24**, 429–432 (1986).
- Spielman, A. I., Brand, J. G. & Wysocki, L. *Chem. Senses* **14**, 841–846 (1989).
- Devereaux, J., Haerberli, P. & Smithies, O. *Nucleic Acids Res.* **12**, 387–395 (1984).
- Ausubel, F. M. *et al. Current Protocols in Molecular Biology* (Wiley, New York, 1989).
- Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning* (Cold Spring Harbor Laboratory, New York, 1989).
- Krieg, P. A. & Melton, D. A. *Meth. Enzym.* **155**, 397–415 (1987).
- Lugo, D. I., Roberts, J. K. & Pintar, J. E. *Molec. Endocrin.* **3**, 1313–1324 (1989).

ACKNOWLEDGEMENTS. We thank A. Spielman for advice on collecting taste tissue; N. Spickofsky and Z. Hao for technical assistance; F. Margolis for discussions; S. Udenfriend, R. Horn and F. Margolis for comments on the manuscript.

LETTERS TO NATURE

An example of stable chaos in the Solar System

Andrea Milani & Anna M. Nobili*

Gruppo di Meccanica Spaziale, Dipartimento di Matematica, Università di Pisa, Via Buonarroti 2, I-56127 Pisa, Italia
Observatoire de Paris, Meudon, Groupe EUROPA-DAEC, 5 place Jules Janssen, F-92195 Meudon, France

MANY planets have been shown to have chaotic instabilities in their orbital motions, but the long-term significance of this is not fully understood¹. The eccentricity of Mercury, for example, changes by about 25% of its value over 40 times the Lyapunov time^{2,3} (the e-folding time for divergence of nearby orbits), but the orbit of Pluto, in an integration lasting 50 Lyapunov times⁴, shows no significant change. Here we show that the orbit of the near-Jupiter asteroid 522 Helga is chaotic, with an unusually short Lyapunov time of 6,900 yr. We integrate its motion, including perturbations from the outer giant planets, over a period 1,000 times longer than this, and find no significant instability. Chaos in the orbit of 522 Helga is caused by a 7:12 resonance with the orbit of Jupiter, but the size of the chaotic region in phase space is small; stability is ensured because the eccentricity and precession of the orbit are such that it avoids close encounters with Jupiter. Asteroid orbits with larger proper eccentricity would, we suggest, be genuinely unstable, consistent with the sparse asteroid population near Helga. Although Helga is the first clear-cut example of a stable chaotic orbit, we argue that 'stable chaos' may be a rather common feature of Solar System dynamics.

A dynamical system with a positive maximum Lyapunov

exponent is defined to be chaotic. An indication of the maximum Lyapunov exponent can be obtained by monitoring the function

$$\gamma(t) = \log \frac{|v(t)|}{|v_0|} \quad (1)$$

where $v(t)$ is the solution of the equations of relative motion, linearized around the solution under study, with initial conditions $v(0) = v_0$ chosen at random. Figure 1a shows such a function for the orbit of the asteroid 522 Helga. By definition, the maximum Lyapunov exponent is $\lim_{t \rightarrow +\infty} \gamma(t)/t$. For Helga, the best fit over a 7-Myr integration has a slope of $1.45 \times 10^{-4} \text{ yr}^{-1}$, so the Lyapunov time is only $\sim 6,900 \text{ yr}$. There is no way to know whether such a rate would be maintained for all time, but the divergence of two nearby orbits by a factor $\sim \exp(1,000)$ certainly appears as a very strong form of local instability. We have calculated the reference orbit of 522 Helga, as well as the variational equations, within a fairly complete physical model in three dimensions. It accounts for the full gravitational perturbations of the outer planets Jupiter, Saturn, Uranus and Neptune on the massless asteroid. Perturbations by the inner planets are not important for outer-belt asteroids such as Helga, so they have been neglected. All initial conditions have, however, been referred to the centre of mass of the inner Solar System, to decrease the secular effects of the neglected perturbations. The observed chaotic behaviour is therefore not an abstract property of some simplified model, but is intrinsic to the orbit of the real asteroid. In fact, simpler models give qualitatively similar results. The actual value of the Lyapunov exponent is, however, sensitive to changes in both the physical model and the initial conditions. Values differing by 20% result from relative changes of only 10^{-5} in the initial elements. On the other hand, the calculation of the Lyapunov exponent over a finite timespan is only an approximation. Calculations over different timespans confirm the chaotic character of the orbits,

* To whom correspondence should be addressed at the Observatoire de Paris.

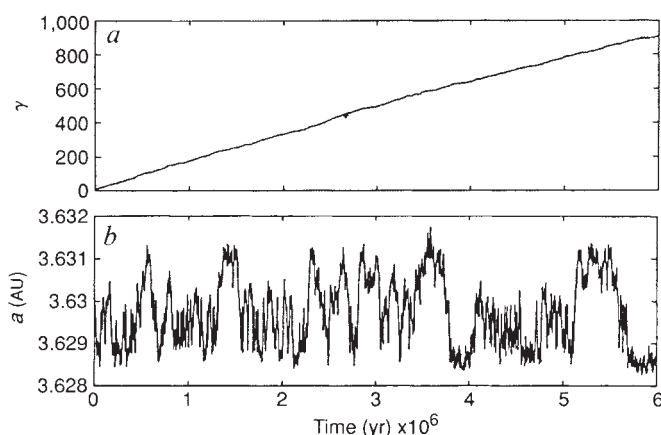


FIG. 1 *a*, Asteroid 522 Helga. Logarithm of the divergence of two nearby orbits, as calculated from the linearized equations of the relative motion, plotted against time for a timespan of 7 Myr. *b*, The semimajor axis *a* of the asteroid 522 Helga (in astronomic units) as a function of time over 7 Myr. Short periodic perturbations have been removed by digital filtering.

but the exact value of the estimated exponent is not reproducible.

The orbital elements of 522 Helga do not show signs of large-scale instabilities. The comparatively large variations in semimajor axis (Fig. 1*b*) seem to be associated with the orbit getting in and out of the 7:12 resonance between the mean motion of the asteroid and the mean motion of Jupiter. The argument $7\lambda - 12\lambda_J$ (where λ is the mean longitude of the asteroid, and λ_J that of Jupiter) is a slow variable, completing only ~ 830 revolutions in 7 Myr. Periods of ostensible libration alternate with periods of circulation, which is typical of a chaotic situation. Because all secular resonances act over much longer timescales, only a mean motion resonance can be responsible for chaos acting over a timescale as short as 6,900 yr.

With a semimajor axis between 3.628 and 3.632 AU, an eccentricity between zero and 0.094 and a small inclination ($< 3.9^\circ$), Helga could in principle get as close as ~ 0.9 AU to Jupiter. So small an approach distance would make the apparent macroscopic stability over the 7-Myr integration time difficult to understand. Instead, we find that Helga is never closer to Jupiter than 1.31 AU, and the reason can be understood from Fig. 2*a*, where the distance from the origin is the asteroid's osculating eccentricity at that time, and the polar angle is the angle $\varpi - \varpi_J$ between the perihelion of the asteroid and that of Jupiter. In a more-simplified model in which the perihelion of Jupiter is fixed, the plot would appear like a circular annulus with a fixed centre⁵. The distance of the central point from the origin is the eccentricity e_F forced by Jupiter at the asteroid location⁶, and the mean radius of the annulus is the proper eccentricity e_P of the asteroid. Although the osculating eccentricity has rather large variations (from a minimum of $|e_F - e_P|$ to a maximum of $e_F + e_P$, depending on the value of the angle $\varpi - \varpi_J$), the relevant physical quantity is the proper eccentricity which tells us about the origin and evolution of the orbit.

Figure 2*a* shows a moving annulus because of the variation of e_F , which, being proportional to the eccentricity of Jupiter, feels the perturbations of the other planets (mainly that of Saturn) on Jupiter. Consequently, the polar angle shows both libration around 0° and circulation, depending on whether e_F is large or small. Libration around 0° means that the perihelion of Helga and the perihelion of Jupiter are locked to one another, thus ensuring that the asteroid is never at aphelion in conjunction with Jupiter's perihelion (which would give the closest possible approach between the two). This mechanism of dynamical protection is known as ϖ libration. Instead, as the polar angle rotates the two pericentres can become 180° apart, and the asteroid can be at its aphelion with Jupiter at perihelion in the same direction. In this configuration, however, the osculating eccentricity is always near its minimum, and the eccentricity of Jupiter is also near its minimum. The mechanism is known as e - ϖ coupling. Figure 2*a* shows that either the ϖ libration or the e - ϖ coupling is active throughout the 7-Myr timespan of

the numerical integration. As a result, the minimum distance between Helga and Jupiter oscillates between 1.31 AU and 1.57 AU, explaining the observed macroscopic stability.

Figure 2*a* contains further important information, as it allows us to compute the proper eccentricity. This can be determined numerically by transforming the moving annulus into a fixed

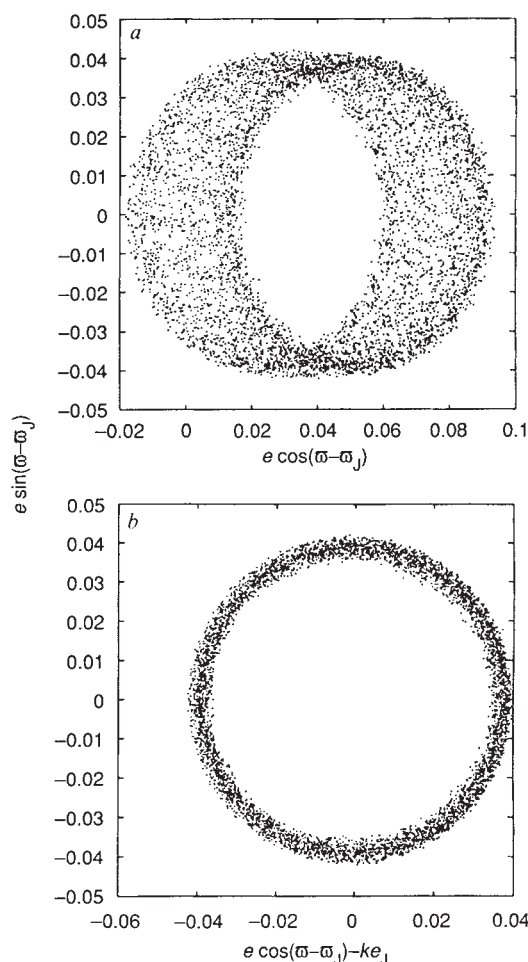


FIG. 2 *a*, 522 Helga. The variable $e \sin(\varpi - \varpi_J)$ is plotted against $e \cos(\varpi - \varpi_J)$ over 7 Myr. The distance from the origin is the osculating eccentricity of the asteroid at the time; the polar angle is the angle between the pericentre of Helga and that of Jupiter. One data point has been plotted for every 1,000 yr. *b*, Same as *a*, after removing a time-dependent forced eccentricity.

one by subtracting from the variable $e \cos(\varpi - \varpi_J)$ a multiple ke_J of Jupiter's eccentricity⁷. The constant k is determined by a fit to the data of Fig. 2a in such a way that it is transformed into a new distribution centred at the origin (Fig. 2b). The radius of the annulus in Fig. 2b is then the proper eccentricity. Obviously this is not constant, but its average value is 0.039, with a standard deviation of 0.0015. The proper elements of main-belt asteroids can be calculated analytically⁸ with comparable stability, but this is not possible so close to a mean-motion resonance. Thus the proper eccentricity is stable within a narrow range, as is the proper inclination, so the chaotic region is very small. The small value of Helga's e_p , with $e_p < e_F$ most of the time, is the true explanation for its long-term macroscopic stability, as both the protection mechanisms that are effective in increasing the minimum close-approach distance to Jupiter are a direct consequence of the relations $e_p < e_F$ (ϖ libration), $e_p \geq e_F$ ($e - \varpi$ coupling). All objects in the region with an original proper eccentricity significantly larger than the forced one must have been ejected, as suggested by the fact that the asteroid population in the region around Helga is very sparse⁹. It is worth stressing that the protection mechanisms effective for Helga have nothing to do with the 7:12 resonance responsible for the observed chaotic behaviour.

The most spectacular examples of chaos in the Solar System are those in which the chaotic domain contains a region of very close approaches (even the singularity of collision) and chaos does explain the absence of real objects¹⁰⁻¹³. The planets themselves have chaotic orbits¹⁴⁻¹⁶, but the significance of chaos is different. For the major outer planets the chaotic region is very thin (10^{-4} in eccentricity)¹⁵. In contrast, chaos causes an absolute change of 0.05 in the eccentricity of Mercury over only 40 times the Lyapunov time^{2,3}, and this is by no means a weak instability. The orbit of Pluto has now been calculated for 50 times the Lyapunov time⁴; the chaotic character of the orbit is confirmed but still it shows no significant instability. The relatively short Lyapunov time was taken as an indication of robust chaos¹⁴, but the chaotic region of the high-order secular resonance which is the best candidate for originating Pluto's chaos seems to be small¹⁷. Examples of chaos are being found throughout the Solar System, but their true meaning is not yet understood¹. Our results show a clear-cut example in which the Lyapunov exponent has little significance. Helga provides an example of stable chaos: the causes of both its strong chaos and its stability are understood, within a realistic physical model over 1,000 times the Lyapunov time. Although a rigorous theory for the occurrence of stable chaos is not available, it can be argued that if the resonance responsible for the onset of chaos is not the main protection mechanism, then the chaotic domain does not contain a region where the perturbations are significantly increased and therefore macroscopic instability should not be expected. It is therefore likely that stable chaos occurs in many other problems of Solar System dynamics. □

Received 2 December 1991; accepted 14 April 1992.

1. Nobili, A. M. & Burns, J. A. *Science* **244**, 1425 (1989).
2. Laskar, J. *Icarus* **88**, 266-291 (1990).
3. Laskar, J. in *Chaos et Déterminisme* (eds Dahan, A., Chabert, J. L. & Chemla, K.) (Seuil, Paris, in the press).
4. Wisdom, J. & Holman, M. *Astr. J.* **102**, 1528-1538 (1991).
5. Milani, A. & Nobili, A. M. *Celest. Mech.* **34**, 343-355 (1984).
6. Brouwer, D. & Clemence, G. M. *Methods of Celestial Mechanics* Ch. XVI (Academic, New York, 1961).
7. Schubart, J. *Celest. Mech.* **43**, 309-317 (1988).
8. Milani, A. & Knežević, Z. *Celest. Mech.* **49**, 347-411 (1990).
9. Nobili, A. M. in *Asteroids II* (eds Binzel, R. P., Geheerls, T. & Matthews, M. S.) 862-879 (University of Arizona Press, 1989).
10. Bevilacqua, R., Menchi, O., Milani, A., Nobili, A. M. & Farinella, P. *Moon Planets* **22**, 141-152 (1980).
11. Wisdom, J. *Icarus* **56**, 51-74 (1983).
12. Wisdom, J. *Nature* **315**, 731-733 (1985).
13. Milani, A. & Nobili, A. M. *Astr. Astrophys.* **144**, 261-274 (1985).
14. Sussman, G. J. & Wisdom, J. *Science* **241**, 433-437 (1988).
15. Nobili, A. M., Milani, A. & Carpino, M. *Astr. Astrophys.* **210**, 313-336 (1989).
16. Laskar, J. *Nature* **338**, 237-238 (1989).
17. Milani, A., Nobili, A. M. & Carpino, M. *Icarus* **82**, 200-217 (1989).

ACKNOWLEDGEMENTS. The authors are on leave at the Observatoire de Paris with the ESA G. Colombo fellowship (A.M.) and a NATO senior fellowship (A.M.N.).

Bulk synthesis and selective exchange of strontium ions in $\text{Na}_4\text{Mg}_6\text{Al}_4\text{Si}_4\text{O}_{20}\text{F}_4$ mica

William J. Paulus*, Sridhar Komarneni* & Rustum Roy

Materials Research Laboratory, The Pennsylvania State University, University Park, Pennsylvania 16802, USA

* W.J.P. is currently at AC Rochester Division of General Motors Corporation, Department 32-84, 1300 North Dort Highway, Flint, Michigan 48556, USA. S.K. is also with the Department of Agronomy

ONE of the principal long-term problems associated with nuclear fission reactions is the generation of large amounts of radioactive strontium isotopes, particularly ^{90}Sr . An effective method for the removal and entrapment of strontium from contaminated environments is therefore desirable to prevent potential health problems associated with radionuclide ingestion¹⁻⁵. In this context, inorganic cation exchangers have been useful by virtue of their capacity for structure-specific ion exchange⁶⁻¹². We report here that 'Na-4-mica'¹³⁻¹⁵, a highly charged sodium fluorophlogopite mica ($\text{Na}_4\text{Mg}_6\text{Al}_4\text{Si}_4\text{O}_{20}\text{F}_4$), can be used for the selective removal of strontium ions from solution and for strontium immobilization at room temperature. We also report a novel method for bulk synthesis of 'Na-4-mica' as a pure phase. This material is unique among highly charged or brittle micas because of its ability to undergo a hydration reaction, which allows the material to be used as a highly selective strontium-ion sieve. This or similar compounds could find potential application in the decontamination and disposal of nuclear waste.

Environments containing the radionuclide ^{137}Cs can be successfully decontaminated through the use of a highly selective caesium ion sieve made by topotactic leaching of a naturally occurring phlogopite mica^{16,17}. This material was found to be far superior to other naturally occurring and synthetic caesium-exchange materials in both its caesium selectivity and its ability to immobilize caesium permanently after exchange through chemical bonding, ultimately forming a stable crystalline waste

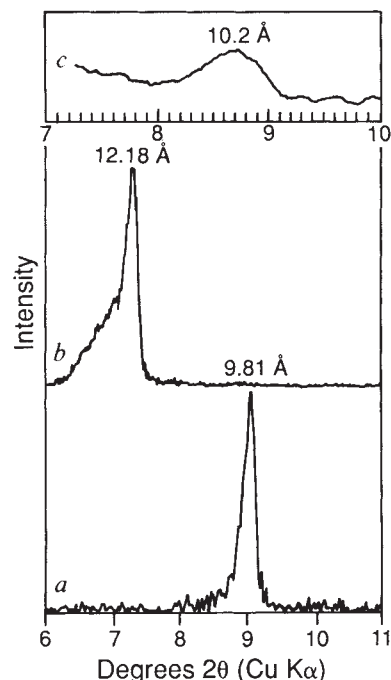


FIG. 1 X-ray diffractograms of Na-4-mica. a, anhydrous form; b, hydrated form; c, collapsed form after exchange with Sr^{2+} ions.

at room temperature. A logical progression was to seek out a similar micaceous material that would work in the same way with ^{90}Sr . We recognized that a possible candidate was Na-4-mica, which was first synthesized in minute quantities (~ 20 mg) during experiments on the synthesis of mica-type silicates from fluoride melts¹³⁻¹⁵. Although Na-4-mica could be synthesized only in small quantities in the presence of large amounts of impurity phases, it has the unique advantage among brittle micas, of becoming hydrated on contact with water or even in moist air at ambient conditions¹⁵. This behaviour arises from its unusually large number of interlayer cations and a subsequent offset layer stacking mode, allowing the structure to expand from a dehydrated 9.81 Å to a hydrated 12.18 Å *c*-axis spacing to achieve a more thermodynamically stable interlayer structure^{15,18}. Once hydrated, the Na-4-mica synthesized by Gregorkiewicz *et al.* was shown to undergo limited topotactic cation exchange with solutions containing strontium, in that the exchange reaction proceeded from the edges of the individual crystallites for a few micrometres into the interior of the mica flakes until further exchange was arrested by layer collapse¹⁵. This behaviour was, in effect, identical to that shown by the caesium-selective ion sieve^{16,17}. If an effective method of synthesis could be developed which could manufacture large amounts of phase-pure Na-4-mica flakes with a small and uniform particle size, we hoped to maximize this mica's strontium-ion exchange capacity and develop an effective strontium-ion sieve which could be used to isolate radioactive strontium.

Our synthesis method, which can be easily scaled up or down as laboratory equipment and facilities allow, used a sol-gel-derived precursor powder dissolved in a molten NaF flux and is briefly outlined as follows¹⁹. A monophasic gel powder was prepared by separately dissolving $\text{Al}(\text{NO}_3)_3 \cdot 9\text{H}_2\text{O}$ and $\text{Mg}(\text{NO}_3)_2 \cdot 6\text{H}_2\text{O}$ in absolute ethanol and then combining and mixing the two solutions for an hour with stirring. A known quantity of tetraethylorthosilicate, $\text{Si}(\text{OC}_2\text{H}_5)_4$, was added to obtain a stoichiometric composition of the mixture, which when reduced to pure oxides was exactly $3\text{MgO} \cdot \text{Al}_2\text{O}_3 \cdot 2\text{SiO}_2$. The resulting sol was stirred for three hours to ensure homogeneity, then tightly covered and placed in a 60 °C oven to promote hydrolysis, condensation and polymerization reactions for three days. The water of hydration of Al and Mg nitrates is sufficient for hydrolysis and therefore no water was added. This resulted in a single-phase gel which was then dried at 100 °C in an oven, calcined at 475 °C for 12 hours, ground to pass through a -325 mesh screen and intimately mixed with an equal mass of -325-mesh crystalline NaF. The mixture was sintered for 18 hours at 890 °C in a platinum vessel in an air-atmosphere furnace. This sintering temperature was chosen because thermal analysis showed that the monophasic gel begins to dissolve into the NaF at just below this point, forming an evenly mixed solution devoid

of large concentration gradients. Furthermore, at this temperature tiny flakes of Na-4-mica, with the ideal formula $\text{Na}_4\text{Mg}_6\text{Al}_4\text{Si}_4\text{O}_{20}\text{F}_4$, nucleate from the melt but do not grow past a particle size of $\sim 1\text{--}5$ µm because of the relatively low temperature. After cooling at furnace rate, the reaction products were washed in deionized water several times to remove the remaining NaF. The remaining low percentage of impurity phases that are not water soluble, which included various amorphous phases and insoluble fluoride salts, were then removed with repeated washings with a saturated boric acid solution. X-ray diffraction and scanning electron microscopy were used to confirm the phase purity of the Na-4-mica, which had a composition very close to the ideal formula $\text{Na}_4\text{Mg}_6\text{Al}_4\text{Si}_4\text{O}_{20}\text{F}_4$ and a *d*(001) spacing of 9.81 Å in the anhydrous form (Fig. 1a). Hydration of Na-4-mica evidently took place as easily as it did in Na-4-mica flakes^{14,15}, as the washing process caused the material to expand to a *d*(001) spacing of 12.18 Å (Fig. 1b).

The hydrated form of Na-4-mica ($\text{Na}_4\text{Mg}_6\text{Al}_4\text{Si}_4\text{O}_{20}\text{F}_4 \cdot x\text{H}_2\text{O}$) was then used in all of the following strontium-exchange experiments. We determined a Sr^{2+} exchange isotherm by equilibrating 0.0200 g of Na-4-mica in 25.0 ml of ultrapure SrCl_2 solutions having normalities between 0.0006 and 0.0037N for 28 days. The solid and solution phases were separated by centrifugation after equilibration, after which the solutions were analysed for Sr^{2+} by atomic emission spectroscopy and the solid phase was characterized by powder X-ray diffraction. We also explored the selective strontium-exchange properties by generating $\text{Na}^+/\text{Sr}^{2+}$ exchange equilibrium with constant amounts of Na-4-mica and constant total solution normalities but with different ratios of Sr^{2+} and Na^+ cations. As before, the solid and solution phases were separated and analysed after equilibration for 28 days. We also determined whether there was selective strontium uptake in the presence of extremely large excess quantities of Na^+ by equilibrating 0.0200 g Na-4-mica for 24 h in 25.0 ml of 0.50N NaCl solution also containing 0.0001N SrCl_2 , which amounted to a solution having a $[\text{Na}^+]/[\text{Sr}^{2+}]$ equivalent ratio equal to 5,000:1. After equilibration, the solid and solution phases were separated and analysed as described above. The selective strontium uptake under these conditions can then be expressed as a distribution coefficient, K_d , defined as the ratio of the amount of strontium sorbed per gram of solid to the amount of strontium remaining per millilitre of solution²⁰. In this case, $K_d > 1 \text{ ml g}^{-1}$ shows preference for strontium. All of the data points in the figures represent the average of triplicate runs with a mean variation of less than $\pm 3\%$.

The strontium-exchange isotherm of Na-4-mica in the presence of pure SrCl_2 solutions (Fig. 2) illustrates that a steady state was attained at a strontium loading of 203 mequiv per 100 g in the presence of Na^+ being released from the interlayers during equilibration. This represents $\sim 44\%$ of the theoretical exchange capacity of Na-4-mica, 462 mequiv per 100 g, indicating that the strontium exchange was incomplete. This incomplete exchange can be explained by the fact that the interlayer spacing collapsed considerably when about half of the exchange sites were occupied by strontium. This was confirmed by powder X-ray diffraction (Fig. 1c) which showed that the *c*-axis spacing decreased from 12.18 to 10.2 Å. The collapse can be explained by dehydration because of the extremely high charge density of the layers coupled with the relatively low hydration energy of the strontium ions²¹⁻²³, and it not only prevents any further exchange of Sr^{2+} ions from the solution for Na^+ in the highly charged Na-4-mica, but also effectively traps the strontium ions that entered the structure between the collapsed layers, resulting in irreversible fixation.

Figure 3 illustrates the equilibrium sodium/strontium exchange isotherm of Na-4-mica. The diagonal line shows equal preference for both Na^+ and Sr^{2+} . Initially, the strontium-exchange isotherm falls above the diagonal line, indicating that Sr^{2+} is highly preferred over Na^+ (ref. 24). The isotherm quickly flattens out, however, and this coupled with the fact that its

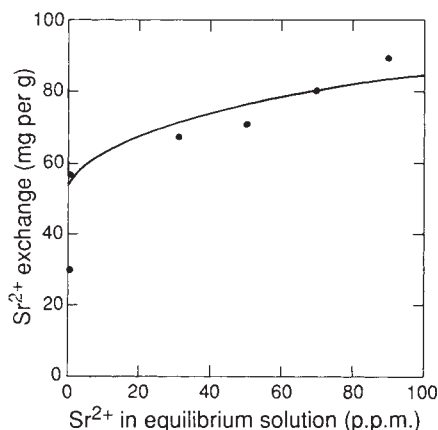


FIG. 2 Ion exchange isotherm of Na-4-mica in the presence of ultrapure SrCl_2 solutions of varying concentrations.

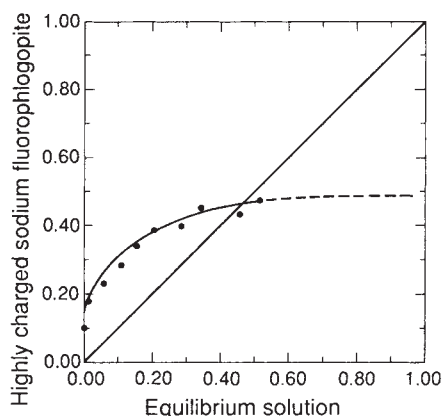


FIG. 3 Selective ion exchange isotherm for Na-4-mica in the presence of mixed ultrapure NaCl/SrCl₂ solutions. Both axes show equivalent fraction of Sr²⁺.

strontium exchange capacity, 216 mequiv per 100 g, is 47% of the theoretical exchange capacity of Na-4-mica indicates that the strontium exchange does not go to completion, because of the interlayer collapse and strontium entrapment described above. Strontium preference over sodium in highly charged Na-4-mica was, therefore extremely high, as indicated by the initial data points of the isotherm falling well above the diagonal line. This extreme preference for strontium over sodium was also shown by the experiment to determine K_d , which yielded a value of 4,720 ml g⁻¹. A further indication of strontium preference is that the total strontium exchange capacity of Na-4-mica is roughly the same whether sodium is present (216 mequiv strontium per 100 g mica) or absent (203 mequiv strontium per 100 g mica) in the equilibrating solution.

Our experiments show that phase pure Na-4-mica synthesised in this way is very useful for strontium exchange. The fine, uniform particle size allows for the number of available exchange sites to be maximized, and the interlayer dimensions and charge density of the material are ideal for strontium diffusion and eventual entrapment. Collapse of the interlayer region achieves effective strontium fixation, as the strontium becomes trapped into a more conventional trioctahedral brittle mica structure similar to the barium interlayered mica, kinoshitalite, Ba₂(Mg, Mn, Al)₆Si₄Al₄O₂₀(OH, F)₄²⁵, which is an extremely stable phase because the high coulombic forces that span the interlayer region effectively hold it very tightly closed¹⁸. We have found that other common divalent ions such as Mg²⁺ and Ca²⁺, which have larger hydrated radii than Sr²⁺, are not selectively trapped between the layers, adding to the strontium selectivity of highly charged sodium fluorophlogopite. □

21. Sawhney, B. *Soil Sci. Soc. Am. Proc.* **33**, 42-46 (1969).
22. Shainberg, I. & Kemper, W. *Soil Sci. Soc. Am. Proc.* **30**, 707-713 (1966).
23. Komarneni, S. & Roy, R. *Clay Miner.* **21**, 125-131 (1986).
24. Breck, D. *Zeolite Molecular Sieves, Structure, Chemistry, and Use* (Wiley, New York, 1974).
25. Guggenheim, S. in *Reviews of Mineralogy*, Vol. 13, *Micas* (ed. Bailey, S. W.) (Mineralogical Society of America-Bookcrafters, Inc., Chelsea, Michigan, 1984).

ACKNOWLEDGEMENTS. This research was supported by the U.S. Department of Energy.

Alteration of tektite to form weathering products

J. J. Mazer*, J. K. Bates*, J. P. Bradley†‡, C. R. Bradley* & C. M. Stevenson§

* Argonne National Laboratory, 9700 South Cass Avenue, Argonne, Illinois 60439, USA

† McCrone Associates Incorporated, 850 Pasquelli Drive, Westmont, Illinois 60559, USA

§ Archaeological Services Consultants, PO Box 02095, Columbus, Ohio 43202, USA

RECENT use of tektites as evidence for a bolide impact at the Cretaceous/Tertiary (K/T) boundary has focused attention on their long-term stability¹⁻⁵. It was proposed in these studies that residual clay features with the spherical tektite morphology result from *in situ* alteration of the original glassy material. By contrast, examination of tektite alteration as an analogue for the long-term degradation of nuclear waste glass has revealed no evidence of alteration, hydration or devitrification either for samples found in nature or for those reacted in the laboratory⁶⁻⁹: no residual clay minerals were observed, and therefore the glass was interpreted as having reacted by a complete dissolution or etching process¹⁰⁻¹². Here we show that these apparently incongruent observations can be reconciled through understanding the relationship between the environment in which the glass reacts and the chemical processes that control the reaction rate. We have examined both natural and experimental alteration of tektites and have found that, under conditions of restricted water contact, tektite reaction is dominated by water diffusion and *in situ* hydrolysis of the glass structure, followed by restructuring of the silicate network to form clays. Over time, the effective rate for these processes is lower than that for etching. Thus alteration of tektites to clays, as observed at the K/T boundary, can proceed only under conditions of limited water contact.

In conditions of submarine and high-dilution water submersion or laboratory-induced leaching, tektites are etched, all elements in the outer glass structure dissolving into solution stoichiometrically at a rate of $\sim 7 \times 10^6 \mu\text{m d}^{-1}$ at 25 °C (ref. 6). But when obsidian and nuclear waste glasses are reacted under conditions where the etching process is suppressed, such as water vapour or intermittent water contact, they alter by a combination of water diffusion into the glass, alkali depletion, *in situ* hydrolysis and clay formation. Obsidian, when exposed to vapour, forms a birefringent layer as water diffuses into the glass structure^{13,14}. Alternatively, under the same conditions, nuclear waste glasses form complex clay assemblages^{15,16}. These observations suggest that the resultant products (layers) of glass reaction depend on the local reaction environment and raise the following question: if tektites are reacted under the subaerial conditions described above, will processes other than etching become dominant?

We examined the alteration of a naturally occurring land-based tektite and compared it with the alteration of fresh segments of the tektite reacted under selected laboratory conditions. The tektite chosen was a surface sample of Indochinite tektite from Thailand which had been exposed to restricted intermittent groundwater, where the ratio of glass surface to water (S/V)

Received 3 February; accepted 14 April 1992.

1. Jacobsen, L., Simmons, E. & Block, M. *J. Lab. clin. Med.* **34**, 1640-1655 (1949).
2. Adler, S., Trobaugh, F. & Knospe, W. *J. Lab. clin. Med.* **89**, 592-602 (1977).
3. Sawyer, R., Strausbaugh, P. & Morahan, P. *Lab. Invest.* **46**, 165-170 (1982).
4. Volkman, A., Chang, N., Strausbaugh, P. & Morahan, P. *Lab. Invest.* **49**, 291-298 (1983).
5. Shibata, Y. & Volkman, A. *J. Immunol.* **135**, 3897-3904 (1985).
6. Kerr, J. *Am. Ceram. Soc. Bull.* **38**, 374-376 (1954).
7. Ames, L. L. *Am. Miner.* **46**, 1120-1131 (1961).
8. Hatch, L. & Regan, W. Jr. *Nucleonics* **13**, 27-29 (1955).
9. Sawhney, B. *Soil Sci. Soc. Am. Proc.* **28**, 183-186 (1964).
10. Komarneni, S. & Roy, R. *Nucl. Chem. Waste Management* **2**, 259-264 (1981).
11. Komarneni, S. & Roy, R. *Nature* **299**, 707-708 (1982).
12. Clearfield, A. *Inorganic Ion Exchange Materials* (CRC, Boca Raton, 1982).
13. Gregorkiwitz, M., Alcover, J., Russell-Colom, J. & Serratos, J. in *2eme Reunion des Groupes Europeens d'Argiles*, Strasbourg, 64 (1972).
14. Herrero, C., Gregorkiwitz, M., Sanj, J. & Serratos, J. *Phys. Chem. Minerals* **15**, 84-90 (1987).
15. Gregorkiwitz, M. & Russell-Colom, J. *Am. Miner.* **72**, 515-527 (1987).
16. Komarneni, S. & Roy, R. *Science* **239**, 1286-1288 (1988).
17. Komarneni, S. & Roy, R. *U.S. Patent* 4,808,318 (1989).
18. Giese, R. in *Reviews in Mineralogy*, Vol. 13, *Micas* (ed. Bailey, S. W.) (Mineralogical Society of America-Bookcrafters, Chelsea, Michigan, 1984).
19. Paulus, W. thesis, Pennsylvania State Univ. (1990).
20. Tamura, T. *Am. Assoc. Petrol. Geol. Mem.* **18**, 318-330 (1972).

‡ Present address: MVA Inc., 5500 Oakbrook Parkway 200, Norcross, Georgia 30093, USA.

was high (as compared to submarine samples where the S/V ratio is low). We determined the glass composition by acid dissolution and inductively coupled plasma analysis; the major components were (all values in weight per cent) SiO_2 74.4, Al_2O_3 12.17, $\text{Fe}_2\text{O}_{3(\text{total})}$ 5.02, K_2O 2.61, Na_2O 1.32, CaO 1.52, MgO 1.85, TiO_2 0.76 and MnO 0.11. This composition is in good agreement with other analyses of Indochinite tektite^{8,16}. Although it is enriched in Si and K and depleted in Al, Ca and Na relative to Haitian K/T boundary glass, this tektite is similar in composition to Haitian glass² and in the mid-range for reported tektites¹⁷.

The naturally reacted bottle-green glass was first examined by optical microscopy. The extent of reaction was nonuniform, showing a smooth, glassy surface punctuated with etch pits. An obvious glaze covered the surface and was thickest in the pitted regions. Micrometre-sized sections from various regions on the sample surface were carefully removed with a diamond knife, keeping the orientation of the glaze-glass interface intact. The sections were examined with transmission electron microscopy (TEM) using lattice-fringe imaging, selected-area electron microdiffraction (SAED) and energy-dispersive X-ray spectroscopy (EDS) to characterize the phases.

An overview of the structure of the naturally altered tektite surface is shown in Fig. 1. The glaze coverage is discontinuous and comprised of well crystallized material. The glaze forms an interface with the glass $\sim 0.1 \mu\text{m}$ in thickness, where there is a gradual textural transition from crystalline to amorphous material (as determined by electron diffraction). The glaze consists mainly of regions of two distinct lath-bearing clay phases. The dominant phase is an aluminosilicate clay with a 7-Å basal spacing and a SAED pattern consistent with kaolinite¹⁸. The EDS analyses of this mineral show elemental ratios of 1.00 Si: 0.625 Al; 0.125 Mg: 0.094 Fe: 0.63 K. This composition probably indicates a mixture of 1:1 and 2:1 sheet silicates. The presence of Mg, Fe, K and excess Si suggests that if the 7-Å phase is kaolinite, then it is probably mixed with a 2:1 silicate. The second phase has 10-Å basal lattice fringe spacings and a SAED pattern that includes strong 4.55-Å spacings. These features are compatible with glauconite (and other sheet silicates)^{19–21}. We were unable to analyse this mineral by EDS because of its small domain sizes. Within the clays, infrequent occurrences of haematite ($\alpha\text{-Fe}_2\text{O}_3$), α -quartz (SiO_2), ilmenite (FeTiO_3), Fe-bearing rutile (TiO_2)²² and $\text{TiO}_2(\text{B})$ ²³ were identified by EDS and SAED (Fig. 1).

The morphology of the clay-based glaze also indicates how it was formed. The penetration of clay laths into the glass suggests *in situ* hydrolysis of the glass network followed by

restructuring to form clay-rich material. The minor phases found within the clay may form as elements present in the glass are excluded from the clay and nucleate at separate sites²⁴. Although the reaction is probably glass-dominated because of the high S/V , groundwater action may help to separate the clay glaze from the glass, as is observed in some places.

To investigate tektite reaction under more-controlled conditions of limited water contact, we did experiments with polished $1 \text{ cm} \times 1 \text{ cm} \times 1 \text{ mm}$ monoliths of tektite suspended from a stainless steel support by Teflon thread^{13,25}. Enough distilled and deionized water was included inside the sealed stainless steel vessels to provide a 100% relative humidity environment at the test temperature, and the reaction occurred within the thin film of water that sorbs onto the tektite surface at high S/V ratio¹³. Adequate water was available to promote continued reaction, but the small volume of water prevented continuous etching of the glass. Experiments were done at 150, 175, 200 and 225 °C for periods between 3 and 400 days.

The first stage of reaction produces an optically visible birefringent layer (Fig. 2a). No birefringence has been reported previously for altered tektites, and we believe that the layer is analogous to that formed during natural and experimental obsidian hydration. The birefringent layer formed on obsidian is attributed to strain resulting from water diffusing into the glass network^{26,27}. Partial dissociation of the water occurs, but the silicate network remains intact²⁸. The layer can be observed both by optical microscopy and with dark-field imaging using TEM. The layer thickness increases with time t as a function of $t^{1/2}$ (Fig. 3a), consistent with a water-diffusion process. The temperature dependence, shown by an Arrhenius plot (Fig. 3b), yields an activation energy of 91 kJ mol^{-1} and an extrapolated diffusion rate at 25 °C of $1.6 \times 10^{-7} \mu\text{m}^2 \text{ d}^{-1}$.

The birefringent region formed after 22 days at 175 °C consists of two distinct layers with a total thickness of $1.2 \mu\text{m}$ as seen by TEM (Fig. 2b). The thicker inner fraction has a morphology and composition identical to the original glass, and the slight contrast difference seen in dark-field imaging is probably due to water uptake without breakage of the silicate network. The outer fraction (100 nm) shows a sharp interface with the inner fraction and a partial restructuring of the silicate network to form a poorly crystalline phase with a 10-Å basal spacing (Fig. 2c). The fringes are formed throughout the outer layer with random orientations and indicate that the amorphous, yet hydrated, glass transforms into more-stable, crystalline material. Because of the limited regions of crystallinity, the composition of this mineral was not determined.

Additional matrix reaction occurred after 122 days at 200 °C,



FIG. 1 Transmission electron micrograph of a cross-section of the glaze-glass interface region of a natural, subaerially reacted Indochinite tektite. The sample was prepared using ultramicrotomy and shows a glaze consisting of a mixture of layer silicates plus inclusions of haematite (dark regions). Ilmenite, quartz and iron-bearing rutile are also present in other samples. The alteration of glass to secondary phases results from an *in situ* transformation as shown by the gradual transition from amorphous to crystalline regions.

where the entire outer portion of the 4.3- μm hydration layer restructured and formed a layer silicate (Fig. 2d). The outer portion of the layer was 1.2 μm thick and was completely crystalline, allowing reproducible compositional and structural information to be obtained. SAED analyses for this material indicate a 10-Å basal spacing similar to those found on the naturally altered sample. Rare occurrences of 14-Å basal spacings may be attributed to interstratified smectite clays. Compositional EDS analyses give an average formula of $(\text{Na}, \text{K}, \text{Ca})_{0.56}(\text{Fe}_{0.81}\text{Al}_{2.92}\text{Mg}_{0.62}\text{Ti}_{0.09}\text{Mn}_{0.03})_{4.47}[\text{Si}_{7.33}\text{Al}_{0.67}\text{O}_{20}](\text{OH})_4$. This composition is consistent with a smectite mineral.

The development of the reacted layers with time provides insight into the tektite reaction process. Under saturated vapour conditions, the tektite reacts initially by water diffusion, forming a birefringent layer. Next, complete hydrolysis of the silicate network occurs, as shown by the formation of the partially crystalline outer layer. With extended reaction, this outer layer restructures, forming a completely clay-based layer. For the laboratory-reacted sample, the clay in combination with the hydrated inner layer forms the basis of the birefringent layer. For the natural sample, the clay-rich layer or glaze forms the complete layer. The clay-rich layer is the dominant crystalline product in both the natural and laboratory-reacted samples. The natural Indochinite sample shows greater diversity of minor components than the vapour-reacted tektite, possibly because of (1) longer nucleation and growth periods for the natural

sample, (2) interaction with groundwater, which may affect redistribution of elements within the layer, and (3) the difference in reaction temperature.

The clays that have been associated with altered spheroids from different K/T boundary sites are similar in composition and structure to the minerals formed during Indochinite tektite alteration reported above, including kaolinite, glauconite and smectites^{1-5,17,20}. The unusually high concentrations of Si that we find in our analyses of the minerals with 10-Å and 7-Å basal spacings are similar to those that have been reported for K/T boundary clays¹⁷. Additionally, the minor phases identified in the naturally reacted Indochinite match well with the reported mineralogy of the K/T boundary clay.

It is apparent that when water contact is restricted, tektites will alter to form stable secondary phases. The differences in tektite reaction observed for submarine and the land-deposited tektites described here or found at the K/T boundary can be ascribed to different processes dominating the reaction as the natural environment changes. Under high-dilution water submersion at low S/V , etching of the outer glass surface is the dominant reaction process, and no secondary phases or weathering products have been identified. If, however, the reaction occurs at high S/V (water vapour or small volumes of water), the reaction takes place by water diffusion followed by *in situ* hydrolysis and restructuring. Various clay minerals form, depending on exchange with local groundwater. As the reaction

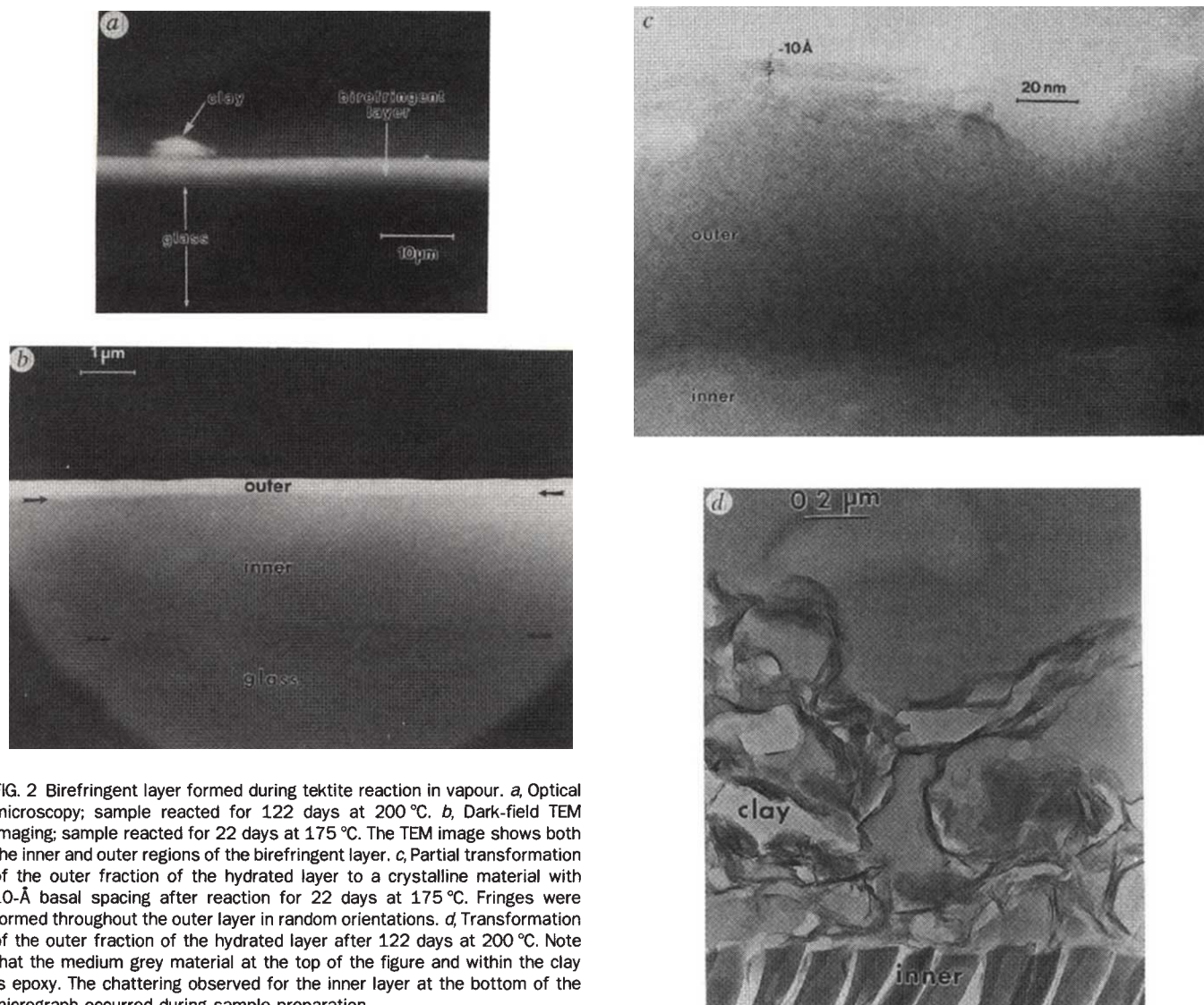


FIG. 2 Birefringent layer formed during tektite reaction in vapour. *a*, Optical microscopy; sample reacted for 122 days at 200 °C. *b*, Dark-field TEM imaging; sample reacted for 22 days at 175 °C. The TEM image shows both the inner and outer regions of the birefringent layer. *c*, Partial transformation of the outer fraction of the hydrated layer to a crystalline material with 10-Å basal spacing after reaction for 22 days at 175 °C. Fringes were formed throughout the outer layer in random orientations. *d*, Transformation of the outer fraction of the hydrated layer after 122 days at 200 °C. Note that the medium grey material at the top of the figure and within the clay is epoxy. The chattering observed for the inner layer at the bottom of the micrograph occurred during sample preparation.

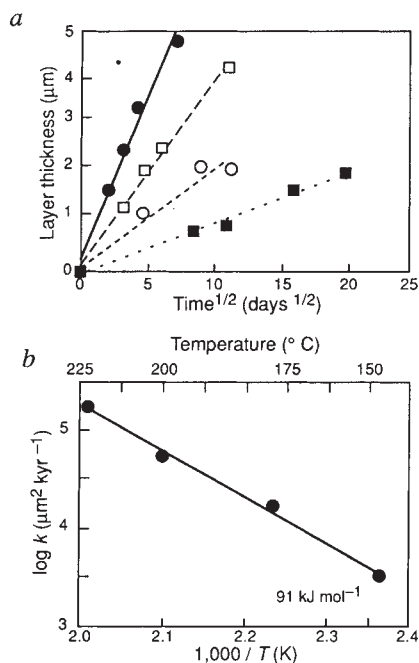


FIG. 3 *a*, Graph of measured hydration rim thickness against the square root of time at (■) 150, (○) 175, (□) 200 and (●) 225 °C for Indochinite tektite reacted in hydration experiments at relative humidity 100%. *b*, Arrhenius plot demonstrating the temperature dependence of the hydration rate constants for tektite hydration at 100% relative humidity. The activation energy is 91 kJ mol⁻¹. Linear regression fits to these data give hydration rate constants of 0.096, 0.208, 0.390 and 0.697 μm day⁻¹, respectively, where the correlation coefficients are greater than 0.98. The errors for the measurements of hydration rate constants fall within the plotted points.

rate for etching exceeds that for diffusion, etching is the dominant reaction process if the reaction conditions are suitable. Our results also suggest that, by examining the structure of reacted natural glass, it may be possible to infer details of the alteration environment. □

Received 3 July 1991; accepted 14 April 1992.

- Montanari, A. *et al. Geology* **11**, 668–671 (1983).
- Sigurdsson, H. *et al. Nature* **349**, 482–487 (1991).
- Margolis, S., Claeys, P. & Kyte, F. *Science* **251**, 1594–1597 (1991).
- Kastner, M., Asaro, F., Alvarez, W. & Alvarez, L. *J. non-cryst. Solids* **67**, 463–464 (1984).
- Izett, G., Dalrymple, G. & Snee, L. *Science* **252**, 1539–1542 (1991).
- Barkatt, A. A. *et al. Geochim. cosmochim. Acta* **48**, 361–371 (1984).
- O'Keefe, J. *Advances in Ceramics*, Vol. 20, *Nuclear Waste Management II*, 689–692 (American Ceramic Society Inc., Westerville, 1986).
- Glass, B. P. *J. non-cryst. Solids* **67**, 333–344 (1984).
- Lutze, W. & Ewing, R. (eds) *Radioactive Waste Forms of the Future* (North Holland, New York, 1988).
- Ewing, R. C. *Scientific Basis for Nuclear Waste Management I*, 57–68 (Plenum, New York, 1979).
- Berger, R. & Ericson, J. in *Recent Advances in Science and Technology of Materials* (ed. Bishay, A.) (Plenum, New York, 1974).
- LaMarche, P. H., Rauch, F. & Lanford, W. A. *J. non-cryst. Solids* **67**, 361–369 (1984).
- Friedman, I. & Long, W. D. *Science* **191**, 347–352 (1976).
- Mazer, J. J., Stevenson, C. M., Ebert, W. L. & Bates, J. K. *Am. Antiq.* **56**, 504–513 (1991).
- Abrajo, T. A., Bates, J. K., Woodland, A. B., Bradley, J. P. & Bourcier, W. L. *Clays Clay Miner.* **38**, 537–548 (1990).
- Advocat, T., Crovisier, J., Vernaz, E., Ehret, G. & Charpentier, H. *Mat. Res. Soc. Symp. Proc.* Vol. 212, 57–64 (Pittsburgh, 1991).
- Kring, D. A. & Boynton, W. V. *Geochim. cosmochim. Acta* **55**, 1737–1742 (1991).
- Jiang, X. & Peacor, D. *Clays Clay Miner.* **39**, 1–13 (1991).
- Brindley, G. W. & Brown, G. in *Crystal Structures of Clay Minerals and their X-Ray Identification* (Mineralogical Society, London, 1980).
- Bornhold, B. D. & Giresse, P. *J. sedim. Petrol.* **55**, 653–664 (1985).
- Montanari, A. *J. sedim. Petrol.* **61**, 315–339 (1991).
- Bursill, L. A. *J. Solid State Chem.* **10**, 72–94 (1974).
- Bankfield, J. F., Veblen, D. R. & Smith, D. *J. Am. Miner.* **76**, 343–353 (1991).
- Bates, J. K., Bradley, J. P. & Bourcier, W. L. *High-Level Radioactive Waste Management: Proc. 2nd ann. int. Conf.*, Vol. 1, 720–727 (American Nuclear Society, La Grange Park, 1991).
- Ebert, W. & Bates, J. K. *Mat. Res. Soc. Symp. Proc.* Vol. 176, 339–346 (Pittsburgh, 1990).
- Brindley, G. W. & Smith, R. *Am. Antiq.* **25**, 476–522 (1960).
- Marshall, R. R. *Geol. Soc. Am. Bull.* **72**, 1493–1520 (1961).
- Lee, R. R., Leich, D. A., Tombrello, T. A., Ericson, J. E. & Friedman, I. *Nature* **250**, 44–47 (1974).

ACKNOWLEDGEMENTS. We thank D. Veblen, G. Ulmer and F. F. Rietmeijer for reviews. This work was supported by the U.S. Department of Energy, Office of Environmental Restoration and Waste Management.

Increased production of cosmogenic ¹⁰Be during the Last Glacial Maximum

Yong Lao*, R. F. Anderson*, W. S. Broecker*, S. E. Trumbore†‡, H. J. Hofmann† & W. Wolfli†

* Lamont-Doherty Geological Observatory of Columbia University, Palisades, New York 10964, USA

† Institut für Mittelenergiephysik, ETH-Honggerberg, CH-8095 Zurich, Switzerland

BERYLLIUM-10 (half-life 1.5 Myr) is produced by spallation of nitrogen and oxygen atoms by cosmic rays in the upper atmosphere. Its production rate is proportional to the flux of cosmic rays, which is modulated by solar activity and the strength of the Earth's magnetic field^{1,2}. Weakening of the magnetic field allows more cosmic rays to impinge on the Earth's atmosphere, thereby increasing ¹⁰Be production. Here we report that the ocean-wide average accumulation rate of ¹⁰Be in Pacific sediments, which reflects the global average production rate of ¹⁰Be (ref. 3), was at least 25% greater during the height of the most recent ice age (~24,000–16,000 yr ago) than during the Holocene (the past 10,000 yr). The higher production rate of ¹⁰Be records the lower intensity of the geomagnetic field during that period and is consistent with the hypothesis developed to explain the younger ¹⁴C ages of fossil corals compared with ages obtained by U–Th dating⁴. These results also point to a more general need to consider variations in production rate in geochronological studies using other cosmogenic nuclides.

The intensity of the Earth's magnetic field varies with time; for example, during the period from ~23,000 to ~15,000 yr BP (before present) its average value may have been ~50% lower than during the past 10,000 yr (ref. 5). Such variations greatly influence the production rates of cosmogenic nuclides such as ¹⁴C and ¹⁰Be. In a study of Barbados fossil corals, Bard *et al.*⁴ found that the ¹⁴C ages of the corals grown during the Last Glacial Maximum (LGM; ~20,000 yr BP) were ~3,500 yr younger than their U/Th ages. They attributed this discrepancy to a ¹⁴C production rate ~40% higher during the LGM, caused by the lower intensity of the geomagnetic field. Palaeointensity data, however, are scarce, and methodological limitations can make them unreliable⁶. The postulated higher production rate of cosmogenic nuclides during lower geomagnetic field intensity must be verified through an independent record such as ¹⁰Be.

The different geochemical cycles of carbon and beryllium require that different approaches be used to evaluate past production rates of ¹⁴C and of ¹⁰Be. After production, ¹⁴C enters a large reservoir of atmospheric CO₂, which is then well mixed, providing a homogeneous atmospheric ¹⁴CO₂ source⁷. In contrast, ¹⁰Be is a particle-reactive nuclide; it becomes associated with aerosols shortly after production and is stripped from the atmosphere on a timescale of months⁸. The deposition of ¹⁰Be onto the Earth's surface varies considerably with time and place and is strongly influenced by local weather conditions^{2,8}. Consequently, the deposition rate of ¹⁰Be at any single site on land (for example in ice cores^{9,10}) may not reliably record changes in its global average production rate. A different approach that integrates ¹⁰Be deposition over a large geographic area must be used. Most of the ¹⁰Be produced in the atmosphere enters the oceans, simply because the oceans cover two-thirds of the Earth's surface. The residence time of dissolved ¹⁰Be in the oceans is long enough (~500–1,000 yr; refs 3, 11) for ocean mixing to smooth out latitudinal gradients in ¹⁰Be deposition from the atmosphere, producing a nearly uniform concentration of ¹⁰Be in the deep waters throughout an ocean basin such as in the

‡ Present address: Department of Geosciences, University of California at Irvine, California 92717, USA.

TABLE 1 Average ratios of $[Pa]_{G/H}$, $[Be]_{G/H}$ and $[Al]_{G/H}$ for cores in the Pacific Ocean

Core	Latitude	Longitude	Water depth (m)	Sedimentation rate (cm kyr ⁻¹)	Holocene* sample age (kyr BP)	Glacial* sample age (kyr BP)	$[Pa]_{G/H}$	$[Be]_{G/H}$	$[Al]_{G/H}$
RC14-121	54° 51' N	170° 41' W	2,532	20.0	1.2-2.1	23.3-28.3	0.38 ± 0.20	1.07 ± 0.39	2.18
V20-122	46° 34' N	161° 41' E	5,563	4.5	2.0-6.0	17.9-23.9	0.95 ± 0.10	1.20 ± 0.10	2.33
RC14-105	39° 41' N	157° 33' E	5,630	6.0	1.8-5.3	18.8-24.8	1.13 ± 0.11	1.67 ± 0.11	1.62
V21-146	37° 41' N	163° 02' E	3,968	3.8	2.8-6.4	19.0-24.0	1.05 ± 0.09	1.64 ± 0.11	1.74
V32-126	35° 19' N	174° 54' E	3,870	2.3	3.3-7.3	17.6-23.8	0.99 ± 0.08	1.43 ± 0.08	1.40
V32-128	36° 27' N	177° 09' E	3,623	1.7	5.2-7.3	17.2-23.9	1.20 ± 0.12	1.95 ± 0.11	2.27
W8709A-1	41° 33' N	131° 57' W	3,680	1.3	5.0-5.6	18.9-21.0	0.64 ± 0.05	1.40 ± 0.07	1.36
W8709A-8	42° 16' N	127° 41' W	3,111	10.0	1.0-4.1	17.4-19.4	0.72 ± 0.09	0.89 ± 0.10	1.44
W8709A-13	42° 07' N	125° 45' W	2,712	20.0	4.7-5.4	19.8-21.6	0.63 ± 0.20	0.81 ± 0.24	1.16
V21-59	20° 55' N	158° 06' W	2,992	0.8	7.1-9.6	14.8-24.3	1.04 ± 0.07	1.33 ± 0.06	1.47
V28-238	01° 01' N	160° 29' E	3,120	1.7	4.8-8.5	21.0-25.3	0.98 ± 0.07	1.50 ± 0.09	ND
RC11-210	01° 49' N	140° 03' W	4,420	1.2	6.0	16.7-19.6	1.02 ± 0.06	1.06 ± 0.06	1.06
V19-28	02° 22' S	84° 39' W	2,720	5.0	5.3	20.6-24.5	0.69 ± 0.06	1.39 ± 0.10	1.13
V19-29	03° 35' S	83° 56' W	3,157	10.0	1.8-4.5	19.8	0.82 ± 0.08	1.18 ± 0.10	1.10
TT154-10	10° 17' S	111° 20' W	3,225	2.0	5.9-6.8	20.1-22.6	0.69 ± 0.04	1.13 ± 0.06	1.47
V19-55	17° 00' S	114° 11' W	3,177	1.2	5.1-7.7	18.8-23.3	0.86 ± 0.05	1.08 ± 0.05	1.48
E17-9	63° 05' S	135° 07' W	4,848	3.6	2.4	18.0	0.65 ± 0.04	0.83 ± 0.05	1.78
RC15-61	40° 37' S	77° 12' W	3,771	3.8	2.0-7.1	18.2-25.4	1.12 ± 0.11	1.02 ± 0.07	1.42
Average†							0.91 ± 0.05	1.25 ± 0.07	1.52

Beryllium-10 was analysed by accelerator mass spectrometer, ^{230}Th and ^{231}Pa by radioactivity measurement, and Al by direct source plasma measurement. $[Pa]_{G/H}$, $[Be]_{G/H}$ and $[Al]_{G/H}$ are the average Pa/Th, Be/Th and Al/Th ratios in the glacial sediments divided by the average Pa/Th, Be/Th and Al/Th ratios in the Holocene sediments, respectively, using equation (2) in the text. The quoted errors for $[Pa]_{G/H}$ and $[Be]_{G/H}$ include all sources of errors propagated throughout the calculation. The approximate uncertainty for analysis of Al in individual samples was ~5%. See ref. 21 for details of methods and sedimentation rates. ND, not determined.

* Two samples from each time period were analysed for ^{10}Be , ^{230}Th and ^{231}Pa nuclides and Al contents for most of the cores. The ages for each core bracket the time interval during which the sediments were deposited. For some cores, only one sample was taken, so only one age is given. The age information is from literature based on ^{18}O , ^{13}C and ^{14}C isotope stratigraphy.

† The uncertainties for the averages of $[Pa]_{G/H}$ and $[Be]_{G/H}$ ratios are the standard error of the mean: for $[Be]_{G/H}$, $n=18$; for $[Pa]_{G/H}$, $n=17$, excluding RC14-121 because the composition of the glacial sediment is very different from that in the Holocene sediment²⁷. For the average of $[Al]_{G/H}$ ratios, $n=17$ because no Al was measured for V28-238.

Pacific¹². If the Pacific Ocean serves as a large homogeneous reservoir of cosmogenic ^{10}Be , just as the atmosphere does for ^{14}C , then a record of the concentration of ^{10}Be in the deep Pacific over the period of interest should indicate the extent to which the global average production rate of ^{10}Be has varied.

No reliable monitor of the concentration of ^{10}Be in sea water has been identified, so the supply of ^{10}Be to the ocean must be evaluated from its rate of accumulation in deep-sea sediments. Evaluating the average accumulation rate of ^{10}Be in marine sediments is, however, complicated by the fact that the intensity of the scavenging processes that remove ^{10}Be from the water column to the sediments varies spatially, leading to a heterogeneous pattern of ^{10}Be accumulation. For example, fluxes of ^{10}Be to sediments near ocean margins, where relatively high biological productivity combined with the supply of eroded continental material greatly enhances scavenging, are more than an order of magnitude greater than fluxes of ^{10}Be to pelagic sediments underlying oligotrophic mid-ocean gyres¹³⁻¹⁶. Given this variability, and the possibility that the regional patterns of scavenging intensity have changed over time, it seems unlikely that a 40% change in ^{10}Be production rate (as is suggested by ^{14}C production rate) can be reliably identified in a marine record of ^{10}Be at a single site, even though the spatial concentration of ^{10}Be in the deep sea is relatively uniform. Two steps are needed to detect a change of this magnitude. First, the ^{10}Be record must be examined at many sites representing a variety of marine environments. Second, to constrain accumulation rates of ^{10}Be more reliably, they should be normalized to a tracer whose source is accurately known.

Naturally occurring uranium-series radionuclides provide the needed tracers. Thorium-230 (half-life 75,200 yr) and ^{231}Pa (half-life 32,500 yr) are produced in the ocean by decay of dissolved ^{234}U and ^{235}U , respectively. Thorium-230 is an extremely particle-reactive nuclide, such that it is removed from sea water on a timescale (decades¹⁷) much less than its radioactive half-life. Consequently, the flux of ^{230}Th to the sediments accumulated

at the time of interest is equal to its well known production rate. Normalizing ^{231}Pa and ^{10}Be to ^{230}Th therefore enables us to evaluate the accumulation rates of ^{231}Pa and ^{10}Be in the ocean. Evaluating such fluxes by normalizing to ^{230}Th has been discussed extensively^{16,18-20}. Briefly, the flux (F) of a sedimentary component N is estimated from the ratio of the concentration of N , $[N]$, to that of initial unsupported ^{230}Th (that produced by decay of ^{234}U in the water column), $[^{230}\text{Th}]$, in sediment as

$$F(N) = P_{\text{Th}}^* [N] / [^{230}\text{Th}] \quad (1)$$

where P_{Th} , the production rate of ^{230}Th , is directly proportional to water depth. We selected 18 independently dated deep-sea cores²¹ (Table 1) to which we applied this technique.

Like ^{10}Be , ^{231}Pa is preferentially removed from sea water in some ocean-margin regions of high particle flux even though the production rate of ^{231}Pa is uniform throughout the ocean¹⁶. An extensive study of Holocene sediments throughout the Pacific has shown that the patterns of enhanced ^{231}Pa deposition are much like those of enhanced ^{10}Be deposition (relative to ^{230}Th) at ocean-margins¹³. The ^{231}Pa to ^{230}Th production ratio has not changed for the last glacial cycle because of the fixed uranium isotopic composition and constant uranium concentration in the ocean due to the very long residence time of uranium (~200,000 yr; ref. 22). Therefore, the average of $^{231}\text{Pa}/^{230}\text{Th}$ ratios of the glacial sediments in a representative set of cores should be the same as that of the Holocene sediments. We test whether the selection of the 18 cores provides an unbiased representation of the average oceanic flux of ^{10}Be using $^{231}\text{Pa}/^{230}\text{Th}$ ratios in the sediments. For this purpose, we introduce a new parameter.

$$[N]_{G/H} = [N/\text{Th}]_G / [N/\text{Th}]_H \quad (2)$$

where N is the concentration of either initial unsupported ^{231}Pa (that produced by decay of ^{235}U in the water column), or ^{10}Be ; Th refers to the concentration of initial unsupported ^{230}Th ; G and H indicate the average results for samples from the glacial

(~24,000 to ~16,000 yr BP) and the Holocene (the past 10,000 yr), respectively. Because the production rate of ^{230}Th (P_{Th}) is constant in equation (1), the $[N]_{\text{G/H}}$ ratio is indicative of a change in the flux of ^{231}Pa or ^{10}Be from the glacial to the Holocene.

At any individual site, the deviation of $[\text{Pa}]_{\text{G/H}}$ from 1.0 could reflect changes in patterns of scavenging intensity which may result from changes in the pattern of particle flux or in sediment particle composition. But the ocean-wide average $[\text{Pa}]_{\text{G/H}}$ ratio must, as noted above, be equal to 1.0. The average of the $[\text{Pa}]_{\text{G/H}}$ ratios is 0.91 ± 0.05 (Table 1), close to 1.0, indicating that the cores selected here provide a reasonable representation of the average sediment in the Pacific (with respect to deposition of the tracer nuclides). In contrast to $[\text{Pa}]_{\text{G/H}}$, the average $[\text{Be}]_{\text{G/H}}$ ratio is 1.25 ± 0.07 (Table 1), suggesting that the spatially averaged flux of ^{10}Be to the deep Pacific sediments during the glacial was ~25% greater than during the Holocene. For reasons discussed below, this value probably represents a lower limit for an increased global average production rate of ^{10}Be during the glacial.

Increased erosion of continents cannot have supplied the extra ^{10}Be . Flux of ^{10}Be associated with continental dust has been estimated²¹ to be ~5% of the total supply of ^{10}Be to the deep Pacific during the Holocene, assuming the ^{10}Be content of Chinese loess²³ to be representative of that of continental dust deposited over the Pacific. We measured the aluminum content of the sediments and calculated from this the fluxes of aluminosilicate dust to the Pacific during the glacial relative to the Holocene, using equation (2). The $[\text{Al}]_{\text{G/H}}$ ratios indicate that the average flux of continental material to the deep Pacific was ~50% larger during the glacial (Table 1). If the ^{10}Be content of the continental dust did not vary, then this represents a negligible (~2%) increase of the continent-derived ^{10}Be accumulated in the deep Pacific during the glacial compared with that during the Holocene.

Virtually all of the ^{10}Be carried by rivers is removed in estuaries¹⁰, so changes in the riverine flux of ^{10}Be cannot have influenced ^{10}Be accumulation in deep Pacific sediments. Beryllium-10 supplied to the oceans with meltwaters during the period of glacial retreat (from ~14,000 to ~8,000 yr ago; ref. 24) may have contributed ~7% of the total supply of ^{10}Be to the world oceans, if we assume the ^{10}Be concentration in ice sheets on the North American and European continents to be the same as that in today's Greenland ice²⁵. Most of meltwater ^{10}Be would have been supplied to the Atlantic (through confined drainage basin systems) and removed in the estuaries there. Changes in the pattern of ocean circulation are unlikely to have significantly increased the partitioning of ^{10}Be from other oceans into the Pacific¹³. Therefore, the deep Pacific acts as a relatively closed basin with respect to ^{10}Be after its deposition from the atmosphere. We have thus estimated the global average production rate of ^{10}Be during the Holocene ($\sim 1.5 \times 10^6 \text{ atoms cm}^{-2} \text{ yr}^{-1}$) from the average accumulation rate of ^{10}Be in the Holocene sediments in the Pacific¹³.

Because bioturbation tends to smooth out signal differences, the average $[\text{Be}]_{\text{G/H}}$ ratio of 1.25 may be a lower limit because sediment accumulation rates in some of the open-ocean cores are sufficiently low that some component of glacial sediments continues to be retained in the surface mixed layer. Furthermore, as the average $[\text{Pa}]_{\text{G/H}}$ ratio is slightly less than 1.0 (Table 1), the ocean-wide average accumulation rate of ^{231}Pa during the glacial may be slightly underestimated using the cores in this study. The average accumulation rate of ^{10}Be during the glacial may similarly be underestimated because the pattern of scavenging of ^{10}Be is similar to that of ^{231}Pa .

In summary, our measured average $[\text{Be}]_{\text{G/H}}$ ratio of 1.25 is a lower limit for the change in the average accumulation rate of ^{10}Be in the Pacific Ocean; supply of ^{10}Be to the Pacific (therefore the production rate of ^{10}Be) must have been at least 25% larger during the glacial than during the Holocene. The

result of this study is clearly consistent with the finding that ^{14}C production rate may be ~40% higher during the LGM⁴, and supports the hypothesis relating the increased production of ^{14}C to a lower geomagnetic field strength. Implications for ^{14}C dating of climate records, archaeological artefacts and so on are obvious. Exposure age dating of surface rocks and landforms based on *in situ* production rates of cosmogenic nuclides such as ^{10}Be , ^{26}Al and ^{36}Cl (ref. 26) will be affected similarly by fluctuations of the geomagnetic field intensity. Variability of the geomagnetic field intensity further back in time, to ~300,000 yr BP, can be examined using the approach described here, ultimately limited by the precision with which unsupported ^{230}Th can be measured in marine sediments. □

Received 9 March; accepted 1 May 1992.

1. Lal, D. & Peters, B. *Encyclopedia of Physics* 46/2, 551–612 (Springer, New York, 1967).
2. Lal, D. *Geophys. Res. Lett.* **14**, 785–788 (1987).
3. Lao, Y. *et al. Earth planet Sci. Lett.* (in the press).
4. Bard, E., Hamelin, B., Fairbanks, R. G. & Zindler, A. *Nature* **345**, 405–410 (1990).
5. Mazaud, A., Laj, C., Bard, E., Arnold, M. & Tric, E. *Geophys. Res. Lett.* **18**, 1885–1888 (1991).
6. Banerjee, S. K. *From Mantle to Meteorites: A Gsrland of Perspective* (eds Gopalan K. *et al.*, 229–238 (Indian Academy of Sciences, Bangalore, 1989).
7. Libby, W. F. *Radiocarbon Dating* (University of Chicago Press, 1952).
8. Brown, L., Stenslend, G. J., Klein, J. & Middleton, R. *Geochim. cosmochim. Acta* **53**, 135–142 (1989).
9. Yiu, F., Raisbeck, G. M., Bourles, D., Lorus, C. & Barkov, N. I. *Nature* **316**, 616–617 (1985).
10. Raisbeck, G. M. *et al. Nature* **326**, 273–277 (1987).
11. Kusakabe, M. *et al. Earth planet. Sci. Lett.* **102**, 265–276 (1991).
12. Kusakabe, M. *et al. Earth planet. Sci. Lett.* **82**, 231–240 (1987).
13. Mangini, A. *et al. Nucl. Instr. Meth. Phys. Res. B5*, 353–358 (1984).
14. Brown, L., Klein, J. & Middleton, R. *Geochim. cosmochim. Acta* **49**, 153–157 (1985).
15. Eisenhauer, A. *et al. Geological History of the Polar Oceans: Arctic versus Antarctic* (eds Bluel, U. & Thiede, J.) 475–487 (Kluwer, Dordrecht, 1990).
16. Anderson, R. F. *et al. Earth planet. Sci. Lett.* **96**, 287–304 (1990).
17. Anderson, R. F., Bacon, M. P. & Brewer, P. G. *Earth planet. Sci. Lett.* **62**, 7–23 (1983); **66**, 73–90 (1983).
18. Bacon, M. P. & Rosholt, J. N. *Geochim. cosmochim. Acta* **46**, 651–666 (1982).
19. Suman, D. O. & Bacon, M. P. *Deep-Sea Res.* **36**, 869–878 (1989).
20. Francois, R., Bacon, M. P. & Suman, D. O. *Paleoceanography* **5**, 861–877 (1990).
21. Lao, Y. thesis, Columbia Univ. (1991).
22. Cochran, J. K. *Uranium Series Disequilibrium: Applications to Environmental Problems* (eds Ivanovich, M. & Harmon, R. S.) 384–430 (Clarendon, Oxford, 1982).
23. Shen, C. thesis, Univ. of Bern (1986).
24. Fairbanks, R. G. *Nature* **342**, 637–642.
25. Beer, J. *et al. Nature* **331**, 675–679 (1988).
26. Lal, D. A. *Rev. Earth planet. Sci.* **16**, 355–388 (1988).
27. Sancetta, C., Heusser, L., Labeyrie, L., Naidu, A. S. & Robinson, S. W. *Mar. Geol.* **62**, 55–68 (1985).

ACKNOWLEDGEMENTS. We thank R. G. Fairbanks for reviewing an early draft of this manuscript. This research was supported by the US NSF.

Enhanced Cenozoic chemical weathering and the subduction of pelagic carbonate

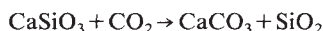
Ken Caldeira

Earth System Science Center and Department of Geosciences, Pennsylvania State University, 248 Deike Building, University Park, Pennsylvania 16802, USA

THE observed trend of increasing oceanic $^{87}\text{Sr}/^{86}\text{Sr}$ ratios during the late Cenozoic led Raymo *et al.*¹ to propose that chemical weathering rates increased at this time as a result of enhanced weatherability of silicate rocks. They suggested that this was due in turn to continental uplift, primarily in the Himalayas and the Andes. Because weathering involves the reaction of silicates with atmospheric carbon dioxide, considerations of changes in weathering rates must take into account the need to balance the global carbon cycle. To maintain this balance on timescales greater than 10^6 yr, enhanced weathering requires an increased flux of CO_2 into the atmosphere². Without such an increased flux, weathering rates could not increase, and one is then forced to search for other explanations for the observed $^{87}\text{Sr}/^{86}\text{Sr}$ trend, such as a changing riverine strontium isotope composition^{3–6}. Here I assume that the strontium isotope record does indeed reflect enhanced weathering in the late Cenozoic, but propose that its cause may have been an

increased CO_2 flux arising from the recycling of pelagic sedimentary carbon at subduction zones. Since the Jurassic, sedimentary carbonate accumulation has shifted from primarily cratonic to primarily pelagic environments⁷⁻¹¹; because sea-floor spreading transports pelagic carbonate to, and recycles it through, metamorphic environments in subduction zones, this shift to pelagic carbonate accumulation would provide the increased CO_2 flux needed to increase weathering rates^{12,13}.

Cenozoic trends in strontium isotopes, pelagic carbonate accumulation and Ge/Si (ref. 14) may indicate enhanced chemical weathering in the late Cenozoic^{1,3,15}. Silicate rock weathering can be schematically represented as^{16,17}



On timescales that are long relative to the $\sim 10^5$ -yr residence time of CO_2 in the atmosphere and oceans¹⁸, such weathering reactions can proceed only as rapidly as CO_2 is supplied to the system. Mountain uplift may increase the rate of weathering reactions for a specified atmospheric CO_2 partial pressure¹. If, however, reactions such as this were to proceed more rapidly than metamorphism and mantle releases supply CO_2 to the atmosphere and ocean, then these CO_2 reservoirs would be drawn down, resulting in a lower atmospheric CO_2 concentration, and chemical weathering rates would return to their value before uplift². An elevated rate of chemical weathering over millions of years therefore requires an increased magmatic or metamorphic CO_2 source.

A simple model can be used to demonstrate the fundamental relationships between the marine Sr-isotope record and the carbon cycle. The primary sources of Sr in the oceans are from continental weathering and mid-ocean-ridge hydrothermal exchange^{3-5,15,19}. Diagenesis of carbonate on the sea floor has little effect on the isotopic composition of marine Sr, because the associated Sr flux is relatively small and its isotopic composition is similar to oceanic strontium⁵. The oceanic $^{87}\text{Sr}/^{86}\text{Sr}$ ratio may be considered to be the time-dependent mean of the $^{87}\text{Sr}/^{86}\text{Sr}$ ratios of the primary sources of oceanic Sr weighted by the magnitude of the respective fluxes⁵, on timescales that are long relative to the 2.5×10^6 -yr residence time of Sr in the oceans³. A schematic steady-state Sr mass-balance for the oceanic $^{87}\text{Sr}/^{86}\text{Sr}$ ratio (R_0^{Sr}), considering only riverine and mid-ocean-ridge sources for oceanic Sr, may be written

$$R_0^{\text{Sr}} = (F_{\text{riv}}^{\text{Sr}} R_{\text{riv}}^{\text{Sr}} + F_{\text{MOR}}^{\text{Sr}} R_{\text{MOR}}^{\text{Sr}}) / (F_{\text{riv}}^{\text{Sr}} + F_{\text{MOR}}^{\text{Sr}}) \quad (1)$$

where F^{Sr} represents Sr fluxes, R^{Sr} represents $^{87}\text{Sr}/^{86}\text{Sr}$ ratios in those fluxes, and subscripts 'riv' and 'MOR' refer to riverine and mid-ocean-ridge hydrothermal sources, respectively.

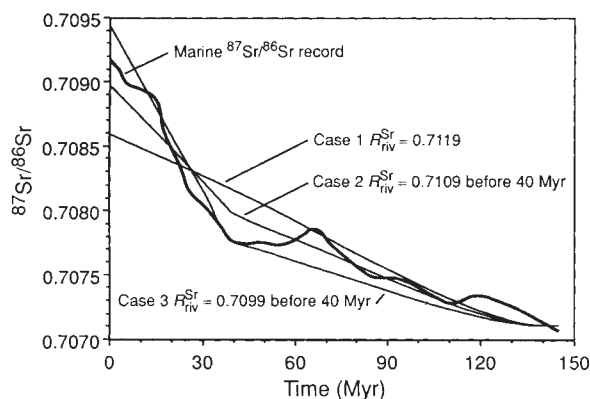


FIG. 1 Model results for the oceanic $^{87}\text{Sr}/^{86}\text{Sr}$ ratio, R_0^{Sr} , for cases 1, 2 and 3 described in the text, compared with the Cretaceous and Cenozoic $^{87}\text{Sr}/^{86}\text{Sr}$ record^{28,29}, smoothed with a 5-Myr moving average. The calculated values of β were 0.30, 0.35 and 0.42 for cases 1, 2 and 3, respectively. Parameter values not specified in the text are: $R_{\text{MOR}}^{\text{Sr}} = 0.7035$, $F_{\text{riv}}^{\text{Sr}} = 33.3 \times 10^9 \text{ mol yr}^{-1}$ (ref. 19) and $F_{\text{MOR}}^{\text{Sr}} = 15.6 \times 10^9 \text{ mol yr}^{-1}$ (ref. 3) for the present day.

Modelling of time-varying sea-floor generation rates is a complication which is unnecessary for the basic argument presented here. If the amount of carbon CO_2 degassing from metamorphic and mantle rocks are roughly proportional to the sea-floor generation and subduction rate¹⁶, then weathering fluxes, including $F_{\text{riv}}^{\text{Sr}}$, will be in rough proportion to the sea-floor generation rate. If hydrothermal exchange rates, including $F_{\text{MOR}}^{\text{Sr}}$, are also roughly proportional to the sea-floor generation rate¹⁶, then the numerator and denominator in equation (1) are both roughly proportional to this rate and therefore R_0^{Sr} is to first order independent of sea-floor generation rate⁶.

The Sr-isotope model represented by equation (1) may be coupled to the carbon cycle by linking the riverine Sr flux ($F_{\text{riv}}^{\text{Sr}}$) to the flux of carbon consumed by continental silicate-rock weathering and subsequent carbonate accumulation ($F_{\text{tot}}^{\text{C}}$). If the silicate-rock weathering rate is directly proportional to the riverine flux of Sr, then this relationship may be represented as

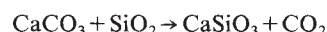
$$F_{\text{tot}}^{\text{C}} = \gamma F_{\text{riv}}^{\text{Sr}} \quad (2)$$

where γ is a constant of proportionality calculated from present-day fluxes. The net CO_2 supply available to weather silicate rocks on the continents is the total metamorphic plus mantle CO_2 supply, less net uptake of sedimentary organic carbon¹⁷ and net CO_2 uptake by the sea floor²⁰. This net CO_2 supply nearly equals the net CO_2 consumption by continental silicate-rock weathering on timescales greater than 10^6 yr (ref. 2). Silicate weathering can therefore increase on long timescales only if there is an increased CO_2 supply available to weather this rock.

The flux of CO_2 available to weather silicate rocks ($F_{\text{tot}}^{\text{C}}$) may be represented as the sum of the flux of CO_2 to the atmosphere or ocean from metamorphic decarbonation of pelagic carbonate ($F_{\text{decarb}}^{\text{C}}$) and the net supply from other sources

$$F_{\text{tot}}^{\text{C}} = F_{\text{decarb}}^{\text{C}} + F_{\text{other}}^{\text{C}} \quad (3)$$

Pelagic carbonate transported to subduction zones may undergo metamorphic decarbonation, which can be schematically represented^{16,17}



This CO_2 may be returned to the atmosphere in back-arc volcanism²¹, in hot springs in regions of metamorphic and magmatic activity²², or from metamorphism in the accretionary prism. At depth in accretionary prisms, temperatures and pressures attain values²³ at which pelagic carbonates can be expected to decarbonate²⁴. Pelagic carbonate is typically transported to metamorphic environments in subduction zones in ~ 50 Myr, whereas the characteristic timescale for the loss of continental carbonate to both weathering and metamorphism is ~ 500 Myr (ref. 8). Increased deep-water carbonate accumulation would therefore increase the rate of carbonate metamorphism¹², and complete transfer of continental carbonate to pelagic environments could increase the global flux of CO_2 from carbonate metamorphism by an order of magnitude¹³.

The pelagic carbonate mass ($M_{\text{pel}}^{\text{C}}$) can be modelled as having a source flux of planktonic carbonate to deep-ocean sediments ($F_{\text{pel}}^{\text{C}}$) and a sink represented by total pelagic carbonate mass divided by a recycling time constant of ~ 50 Myr (ref. 8). It follows that the rate of change of pelagic carbonate mass can be represented as

$$\frac{dM_{\text{pel}}^{\text{C}}}{dt} = F_{\text{pel}}^{\text{C}} - \left(\frac{M_{\text{pel}}^{\text{C}}}{5 \times 10^7 \text{ yr}} \right) \quad (4)$$

The flux of CO_2 to the atmosphere or ocean from metamorphic decarbonation of pelagic carbonate ($F_{\text{decarb}}^{\text{C}}$) is the product of the flux of carbonate transported to subduction zones and the fraction of CO_2 in that flux that is degassed to the ocean or atmosphere (β)

$$F_{\text{decarb}}^{\text{C}} = \left(\frac{M_{\text{pel}}^{\text{C}}}{5 \times 10^7 \text{ yr}} \right) \beta \quad (5)$$

Before the Cretaceous period, most carbonate accumulated in shallow-water environments and pelagic carbonate accumulation was a minor component in the global carbon cycle^{7,8,11}, so that only a small fraction of the CO₂ available to weather silicate rocks could have come from the subduction-zone metamorphism of pelagic carbonate. Through the Cretaceous and Cenozoic there has been a gradual shift from shallow-water to pelagic carbonate accumulation^{8,11}, such that the Recent rate of pelagic carbonate accumulation is the highest^{7,8}, and the Recent area of shallow-water carbonate accumulation the lowest²⁵, of any Phanerozoic epoch. Today, ~60% of carbonate accumulates in pelagic environments⁸.

Wilkinson and Walker⁸ concluded that a considerable flux of planktonic carbonate to deep-ocean sediments ($F_{\text{pel}}^{\text{C}}$) commenced at ~145 Myr ago, and that the total pelagic carbonate accumulation rate has been increasing roughly linearly since then to the present rate (11×10^{12} mol C yr⁻¹, ref. 26). Hence, $F_{\text{pel}}^{\text{C}}$ may be represented

$$F_{\text{pel}}^{\text{C}} = (11 \times 10^{12} \text{ mol C yr}^{-1}) \left(\frac{145 \text{ Myr} - t}{145 \text{ Myr}} \right) \quad (6)$$

where t is time in Myr. Equations (4) and (6) may be solved analytically for $M_{\text{pel}}^{\text{C}}$ with an initial condition of $M_{\text{pel}}^{\text{C}} = 0$ before the onset of significant pelagic carbonate accumulation.

Brass⁴ and Berner⁶ suggested that the marine Sr-isotope record is essentially the record of the relative rates of young and old igneous rock weathering. Francois and Walker²⁷ estimate, however, that the change in these relative weathering rates is less than half that needed to explain the observed increase in marine ⁸⁷Sr/⁸⁶Sr ratios since 160 Myr ago. Holding $R_{\text{MOR}}^{\text{Sr}}$ and the ratio of riverine to mid-ocean-ridge Sr fluxes constant, using equation (1), we find that the values of $R_{\text{riv}}^{\text{Sr}}$ needed to produce the marine ⁸⁷Sr/⁸⁶Sr ratio at 145 Myr ago is 0.7087. This value is less radiogenic than the average ⁸⁷Sr/⁸⁶Sr supplied by the 26 least-radiogenic rivers (out of 86) in the global compilation by Palmer and Edmond¹⁹, representing less than 30% of today's riverine Sr flux. About 25% of that least-radiogenic Sr-flux is from European rivers¹⁹ (Rhine, Danube, Rhone) draining Cretaceous limestones that would not have been present 145 Myr ago.

The global average riverine ⁸⁷Sr/⁸⁶Sr ratio is ~0.7119 (ref. 19). Removing the Ganges, the Brahmaputra and all other rivers with an ⁸⁷Sr/⁸⁶Sr ratio greater than 0.7200 from the global compilation¹⁹ yields a non-radiogenic river average Sr-isotope ratio of 0.7113. It is therefore unlikely that the Himalayan uplift could have caused a shift in riverine Sr-ratios much greater than 0.0006, whereas oceanic Sr-ratios increased by ~0.002 between

145 Myr ago and today²⁸. The 34 least-radiogenic rivers examined by Palmer and Edmond¹⁹ have an average ⁸⁷Sr/⁸⁶Sr ratio of 0.7099. The Cenozoic Sr-isotope record shows a major inflection at ~40 Myr (ref. 29). On the basis of these considerations, I examine three cases: (1) $R_{\text{riv}}^{\text{Sr}} = 0.7119$; (2) $R_{\text{riv}}^{\text{Sr}} = 0.7109$ from 145 Myr to 40 Myr ago, thereafter increases linearly to 0.7119 at 0 Myr ago; (3) $R_{\text{riv}}^{\text{Sr}} = 0.7099$ from 145 Myr to 40 Myr ago, thereafter increases linearly to 0.7119 at 0 Myr.

$F_{\text{other}}^{\text{C}}$ can be calculated from equations (1)–(3) by setting $F_{\text{decarb}}^{\text{C}}$ to zero at 145 Myr, using $R_0^{\text{Sr}} = 0.7071$ (ref. 28). Holding $F_{\text{other}}^{\text{C}}$ constant over time, we can solve equations (1)–(6) for R_0^{Sr} and $F_{\text{riv}}^{\text{Sr}}$ as a function of time. For riverine Sr-isotope cases 1, 2 and 3, constant values of β equal to 0.30, 0.35 and 0.42, respectively, produce the best least-square fit to the Sr isotope curve as compiled by Burke *et al.*²⁸ and Hess *et al.*²⁹ (Fig. 1). These estimates are sensitive to variation in riverine Sr-isotope ratios and carbonate sedimentation. The model calculates present-day metamorphic CO₂ fluxes from subducting pelagic carbonate of 2.2, 2.6 and 3.1×10^{12} mol yr⁻¹ for cases 1, 2 and 3, respectively. This represents roughly one-third to one-half of today's silicate-rock weathering rate¹⁶. The ratio $F_{\text{riv}}^{\text{Sr}}/F_{\text{MOR}}^{\text{Sr}}$ is an indicator of the chemical weathering rate relative to the mid-ocean-ridge hydrothermal exchange rate. The model presented here implies that this ratio is increasing because metamorphic decarbonation enhances chemical weathering without having an appreciable effect on mid-ocean-ridge hydrothermal exchange (Fig. 2).

If some of the increasing flux of pelagic carbonate transported to subduction zones is returned to the atmosphere as metamorphic CO₂, then this should leave an imprint on the marine Sr-record. The above calculations indicate that metamorphic decarbonation of pelagic carbonate could have been sufficient to produce some or all of the Cretaceous/Cenozoic trend in ⁸⁷Sr/⁸⁶Sr ratios measured in marine carbonate, and to produce an increasing trend in Cenozoic chemical weathering rates. Nevertheless, some or all of the variation in the Sr-isotope curve may be due to changes in riverine Sr-isotopic composition^{4–6}. Despite the marked shift in the type of silicate rocks weathered that would be needed to change riverine Sr-isotope composition sufficiently to explain all of the variation in the oceanic Sr-isotope record²⁷, this possibility cannot be ruled out. The Sr-composition of today's rivers may not be representative of Cretaceous rivers⁶ because enhanced volcanism during the Cretaceous, as evidenced by bentonites and volcanogenic sediments^{30,31}, probably led to the development of extensive volcanic areas, including island arcs and rift zones with rapidly weathering, non-radiogenic silicates.

If the marine Sr-isotope record does indicate enhanced Cenozoic chemical weathering relative to hydrothermal exchange rates¹, this enhanced weathering requires an enhanced CO₂ source to balance CO₂ consumed in weathering reactions². Enhanced weatherability from mountain uplift¹ and vegetation changes³², without an increase in the amount of CO₂ available to be consumed in weathering, would decrease atmospheric CO₂ levels without affecting long-term weathering rates. An increased metamorphic CO₂ flux from pelagic carbonate, without enhanced weatherability, would increase both atmospheric CO₂ levels and long-term weathering rates. This suggests that a Cenozoic trend towards decreasing atmospheric CO₂ levels and enhanced long-term weathering rates requires a combination of factors that increase the supply of CO₂ available to weather silicate rock (such as metamorphism of pelagic carbonate) and enhance the weatherability of that rock (such as mountain uplift and vegetation changes). □

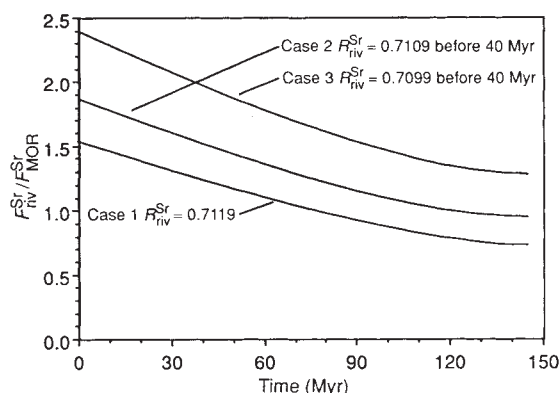


FIG. 2 Model results for $F_{\text{riv}}^{\text{Sr}}/F_{\text{MOR}}^{\text{Sr}}$, an indicator of the global chemical weathering rate relative to the mid-ocean-ridge hydrothermal exchange rate, for cases 1, 2 and 3. If the hydrothermal exchange rate is roughly proportional to the sea-floor generation rate, then $F_{\text{riv}}^{\text{Sr}}/F_{\text{MOR}}^{\text{Sr}}$ times the sea-floor generation rate would be roughly proportional to the silicate-rock weathering rate. Parameter values are as in Fig. 1.

Received 16 December 1991; accepted 7 April 1992.

1. Raymo, M. E., Ruddiman, W. F. & Froelich, P. N. *Geology* **16**, 649–653 (1988).
2. Sundquist, E. T. *Quat. Sci. Rev.* **10**, 283–296 (1990).
3. Hodell, D. A., Mead, G. A. & Mueller, P. A. *Chem. Geol. (Isotop. Geosci. Sec.)* **80**, 291–307 (1990).
4. Brass, G. W. *Geochim. cosmochim. Acta* **40**, 721–730 (1976).

5. Kump, L. R. *Am. J. Sci.* **289**, 390–410 (1989).
6. Berner, R. A. & Rye, D. M. *Am. J. Sci.* **292**, 136–148 (1992).
7. Boss, S. K. & Wilkinson, B. H. *J. Geol.* **99**, 497–513 (1991).
8. Wilkinson, B. H. & Walker, J. G. C. *Am. J. Sci.* **289**, 525–548 (1989).
9. Southam, J. R. & Hay, W. W. *J. geophys. Res.* **82**, 3825–3842 (1977).
10. Davies, T. A. & Worsley, T. R. *Soc. Econ. Paleontol. Miner. Spec. Publ.* **32**, 169–179 (1981).
11. Opdyke, B. N. & Wilkinson, B. H. *Paleoceanography* **3**, 685–703 (1988).
12. Volk, T. *Nature* **337**, 637–640 (1991).
13. Caldeira, K. *Geology* **19**, 204–206 (1991).
14. Shemesh, A., Mortlock, R. A. & Froelich, P. N. *Paleoceanography* **4**, 221–234 (1989).
15. Delaney, M. L. & Boyle, E. A. *Paleoceanography* **3**, 137–156 (1988).
16. Berner, R. A., Lasaga, A. C. & Garrels, R. M. *Am. J. Sci.* **283**, 641–683 (1983).
17. Berner, R. A. *Am. J. Sci.* **291**, 339–375 (1991).
18. Broecker, W. S. & Peng, T.-H. *Tracers in the Sea* (Eldigio, Palisades, New York, 1982).
19. Palmer, M. R. & Edmond, J. M. *Earth planet. Sci. Lett.* **92**, 11–26 (1989).
20. Staudigel, H., Hart, S. R., Schmincke, H.-U. & Smith, B. M. *Geochim. cosmochim. Acta* **53**, 3091–3094 (1989).
21. Marty, B. & Jambon, A. *Earth planet. Sci. Lett.* **83**, 16–26 (1987).
22. Barnes, I., Irwin, W. P. & White, D. E. *Water Resources Investigations, Open-File Rep. 78-0039* (U.S. Geological Survey, Reston, Virginia, 1978).
23. Barr, T. D. & Dahlen, F. A. *J. geophys. Res.* **94**, 3923–3947 (1989).
24. Winkler, H. G. F. *Petrogenesis of Metamorphic Rocks*, 4th Edn (Springer, New York, 1979).
25. Bluth, G. J. S. & Kump, L. R. *Am. J. Sci.* **291**, 284–308 (1991).
26. Berner, E. K. & Berner, R. A. *The Global Water Cycle* (Prentice-Hall, Englewood Cliffs, New Jersey, 1987).
27. Francois, L. M. & Walker, J. C. G. *Am. J. Sci.* **292**, 81–135 (1992).
28. Burke, W. H. et al. *Geology* **10**, 516–519 (1982).
29. Hess, J., Bender, M. L. & Schilling, J.-G. *Science* **231**, 979–984 (1986).
30. Rea, D. K. & Vallier, T. *Geol. Soc. Am. Bull.* **94**, 1430–1437 (1983).
31. Arthur, M. A., Dean, W. E. & Schlanger, S. O. *The Carbon Cycle and Atmospheric CO₂: Natural Variations Archaean to Present* (eds Sundquist, E. T. & Broecker, W. S.) 504–529 (American Geophysical Union, Washington DC, 1985).
32. Volk, T. *Geology* **17**, 107–110 (1989).

ACKNOWLEDGEMENTS. I thank L. R. Kump, M. R. Rampino and J. L. Sloan II for useful discussions, and R. A. Berner and D. L. Des Marais for reviews. This material is based on work supported by the NSF under a grant awarded in 1991.

Helium/heat ratios and deposition temperatures of sulphides from the ocean floor

Grenville Turner & Finlay Stuart

Department of Geology, University of Manchester,
Manchester M13 9PL, UK

CORRELATIONS between excess ^3He concentration and temperature in the ocean above hydrothermal vents have been used^{1,2} to relate the fluxes of heat and noble gases from the upper mantle. Because heat and ^4He are both produced by the decay of uranium and thorium, the observed helium/heat ratios carry information on thermal and mass transport processes within the mantle and at spreading centres. Sulphide and other deposits precipitated on the sea floor trap the hydrothermal solutions as fluid inclusions: we demonstrate here that helium and argon isotopes released by *in vacuo* crushing of sulphides from 21° N on the East Pacific Rise are similar to those in contemporary vent fluids. This observation suggests that it should be possible to extend the study of vent fluid $^3\text{He}/^4\text{He}$ and helium/heat ratios back in time by measuring samples collected from ancient sea-floor deposits.

The intimate link between heat and ^4He production in the Earth has proved useful in constraining large-scale processes operating in the mantle³. Seventy-five per cent of the radioactive heat production in the Earth arises directly from the α -decay of U and Th with a characteristic ^4He /heat production ratio of $3 \times 10^{-8} \text{ cm}^3 \text{ STP J}^{-1}$. Heat is lost predominantly from the Earth through the oceanic crust, much of it in the hydrothermal circulation of sea water in areas of newly formed oceanic crust⁴. But ^4He /heat ratios in ocean-floor hydrothermal fluids vary considerably^{5–11}, ranging from low, but rather uniform values^{6,7} of $\sim 10^{-8} \text{ cm}^3 \text{ STP J}^{-1}$ at mid-ocean-ridge spreading centres to values up to $10^{-6} \text{ cm}^3 \text{ STP J}^{-1}$ at the Loihi hotspot⁹. The observations have important implications for heat and volatile transport mechanisms within the Earth. Furthermore, the recent detection of variations in helium/heat ratios with time, in fluids emanating at 45° N on the Juan de Fuca ridge^{2,11}, indicates that short-term

changes in the volatile and heat flux from the ocean crust may provide a window on magmatic processes.

Hydrothermal vents produce extensive mineral deposition, the most spectacular of these being the black smokers around which are deposited 'chimneys' of iron, copper and zinc sulphides. In common with most hydrothermal minerals, these sulphides trap the emerging helium-rich sea water in fluid inclusions. Because the sulphides are opaque, inclusions so formed cannot be studied directly by optical microscopy, but they can be studied indirectly by means of their noble gas content¹². This suggests that measurements of helium/heat ratios currently made in the ocean could equally well be made on fluids trapped in inclusions. To this end we have analysed the noble gas composition of fluids released by crushing *in vacuo* mineral separates of black smoker and white smoker deposits from the South West and Ocean Bottom Seismometer vent fields at 21° N on the East Pacific Rise collected during Project RISE¹³.

Samples were gently fragmented in a stainless steel mortar and clean mineral separates prepared by hand picking under a binocular microscope to remove tarnished sulphides and contaminating anhydrite. The separates were washed in methanol, then loaded in sample crushers constructed from modified vacuum valves. Before crushing, the samples were baked for 36 hours at 200 °C. The noble gas measurements were interspersed with calibration and blank measurements. In an attempt to estimate blanks directly attributable to the action of the crusher, we also crushed and made measurements on 'inclusion-free' quartz. The blank-corrected He, Ar and Kr abundances, and the isotopic composition of He and Ar, are presented in Table 1, along with sample descriptions. To assess the degree to which the Ar/He ratios may have been altered by preferential diffusive loss of He during the pre-crush baking, three splits of one sample (chalcopyrite ALV 982 R11) were baked for 36, 80 and 156 hours before crushing. No significant or systematic variation in the $^{40}\text{Ar}/^4\text{He}$ ratio was observed.

The $^3\text{He}/^4\text{He}$ ratio of the helium released gives an average of $R/R_a = 7.73 \pm 0.22$ (the measured ratio, R , normalized to the atmospheric ratio, R_a). This is indistinguishable from the values observed in the hydrothermal plumes at 21° N^{6,8} and confirms that minerals formed where the plumes emerge do indeed trap the hydrothermal fluids and retain the helium isotope information.

In every case the isotopic composition of Ar is atmospheric, reflecting its origin in sea water. Figure 1a shows He and Ar data in the form of a plot of $^3\text{He}/^4\text{He}$ against $^{40}\text{Ar}/^4\text{He}$. The distribution of points shows that our measurements do not correspond to a unique composition but lie close to a mixing line between (at least) two endmembers. One of these, M, is a mantle component with $R/R_a = 7.86 \pm 0.12$. Current degassing models¹⁴, measurements on highly vesiculated mid-ocean-ridge basalts¹⁵ and our own data on mantle fluids¹⁶ indicate that $^{40}\text{Ar}/^4\text{He}$ for this component is probably close to 0.6. A second endmember is assumed to be atmospheric noble gases dissolved in the Pacific deep water (including a minor contribution of mantle helium¹⁷) which forms the recharge fluid⁶. The line in Fig. 1a is drawn through the composition of Pacific deep water (PDW; $R/R_a = 1.25$, $^{40}\text{Ar}/^4\text{He} \approx 8,400$)^{17,18}.

Previous studies^{16,19} indicate that atmospheric gases become trapped in minerals exposed to the air for any length of time. This third component (with $R/R_a = 1$, and $^{40}\text{Ar}/^4\text{He} = 1,775$) would increase the Ar/He ratio and obscure the simple two-component interpretation. In our data the presence of an unfractionated air component may be identified by Kr/Ar or Xe/Ar ratios that are consistent with the addition of air to the PDW noble gases (in principle, Ne/Ar could be used, but our Ne measurements are not yet adequate). Table 1 lists the measured Kr abundances using the F -value notation in which $F(\text{Kr})$ is the measured Kr/Ar ratio normalized to the Kr/Ar ratio in the atmosphere²⁰. We have corrected for trapped atmosphere using $F(\text{Kr}) = 1.00$ and $F(\text{Kr}) = 1.98$ (ref. 18) for air and PDW

TABLE 1 Noble gas elemental ratios and isotopes in sulphides from the East Pacific Rise, 21°N

Sample	Description	Weight (mg)	^4He ($10^{-8}\text{cm}^3\text{g}^{-1}$)	R/R_a ($\pm 1\sigma$)*	$^{40}\text{Ar}/^{4}\text{He}$	$^{40}\text{Ar}/^{36}\text{Ar}$ ($\pm 1\sigma$)	$F(^{84}\text{Kr})^\dagger$	$^{40}\text{Ar}/^{4}\text{He}_c$
ALV 982 R11	1–2-mm chalcopyrite from 4-cm-thick interior zone	247	14.4	7.78 ± 0.12	30.6	294.5 ± 0.4	1.38	11.5
ALV 1158-1 no. 3	Anhydrite and minor chalcopyrite from interior 'grey flannel' zone	114	4.4	7.34 ± 0.38	120.8	296.9 ± 0.6	1.55	66.3
ALV 1149-2 no. 15	Pyrite and chalcopyrite from 2-cm-thick rind in chimney interior	228	18.3	7.99 ± 0.18	20.1	296.2 ± 0.7	1.59	11.8
ALV 980 R2	Radially oriented sphalerite from 5-mm-thick outermost zone	158	13.9	7.75 ± 0.12	21.9	295.8 ± 0.7	1.44	9.6
ALV 982 R20	Chalcopyrite and minor pyrite from inner massive sulphide zone	223	20.5	7.85 ± 0.11	67.9	294.3 ± 0.4	1.49	33.3
ALV 979 R5	Fine-grained chalcopyrite from pipe interior	79	15.5	7.65 ± 0.20	44.1	297.1 ± 1.3	1.29	12.6

All data are corrected for procedural blank and mass discrimination. The accuracy of elemental abundance measurements is typically 3% for ^4He and ^{40}Ar , and 10% for ^{84}Kr .

* Measured $^3\text{He}/^4\text{He}$ ratio (R) normalized to the ratio in air (R_a).

† $F(^{84}\text{Kr}) = (^{84}\text{Kr}/^{36}\text{Ar})_{\text{measured}} / (^{84}\text{Kr}/^{36}\text{Ar})_{\text{air}}$.

respectively, and we show the corrected $^{40}\text{Ar}/^4\text{He}$ ratios in Fig. 1b (a correction on the basis of the Xe/Ar ratio is not significantly different but has large errors). The corrected values show a distinct clustering of the sulphide $^{40}\text{Ar}/^4\text{He}$ ratios between 9.6 and 12.6, close to but significantly below the calculated value of 17.7 for contemporary vent water⁸. Ratios of $^{40}\text{Ar}/^4\text{He}$ lower than the vent fluids suggest a higher He flux at the time of sulphide precipitation, but such an interpretation requires independent confirmation.

Ignoring for the moment possible effects produced by phase separation of the hydrothermal fluids^{21,22}, we note that the noble

gas concentration in sea water is fixed within narrow limits. The position of the points along the mantle/seawater mixing line in Fig. 1b is essentially a function of the concentration of the mantle component and can be used to determine that concentration. Furthermore, if the deposition temperature of the minerals containing the fluids can be established, the combination of temperature and concentration of mantle helium can be used to estimate helium/heat ratios, He/Q . It is a simple matter to show that

$$\frac{^3\text{He}_m}{Q} = \frac{^3\text{He}}{^{36}\text{Ar}} \frac{[(^3\text{He}/^4\text{He}) - (^3\text{He}/^4\text{He})_a]}{[(^4\text{Ar}/^4\text{He})_m - (^4\text{Ar}/^4\text{He})_a]} \frac{[^{36}\text{Ar}]_{\text{PDW}}}{C_p \theta} \quad (1)$$

where $[^{36}\text{Ar}]_{\text{PDW}}$ is the concentration of ^{36}Ar in Pacific deep water¹⁸ ($1.26 \times 10^{-6} \text{ cm}^3 \text{ STP g}^{-1}$), C_p the specific heat of sea water¹⁰ ($6.6 \text{ J K}^{-1} \text{ g}^{-1}$) and θ the temperature excess (in $^\circ\text{C}$). Subscripts m and a refer to mantle and atmosphere respectively. Similar expressions can be written relating ^4He to ^{40}Ar , for example

$$\frac{^4\text{He}}{Q} = \frac{^4\text{He}}{^{40}\text{Ar}} \frac{[(^{40}\text{Ar}/^4\text{He})_a - (^{40}\text{Ar}/^4\text{He})]}{[(^{40}\text{Ar}/^4\text{He})_m - (^{40}\text{Ar}/^4\text{He})_a]} \frac{[^{40}\text{Ar}]_{\text{PDW}}}{C_p \theta} \quad (2)$$

The low Ar/He ratios of 21°N basalts²³ compared with that of the deep ocean water means that the mantle contributions to ^{36}Ar and ^{40}Ar can be neglected for all practical applications of equations (1) and (2). We have applied equation (2) to the data in Fig. 1b by assuming that the mean $^{40}\text{Ar}/^4\text{He}$ ratio, 11.4 ± 1.0 , corresponds to the vent exit temperature, 350 $^\circ\text{C}$. Substituting these values in (2), we calculate $^4\text{He}/Q = 1.4 \times 10^{-8} \text{ cm}^3 \text{ STP J}^{-1}$, and $^3\text{He}/Q = 1.5 \times 10^{-13} \text{ cm}^3 \text{ STP J}^{-1}$. The latter is similar to a value of $1 \times 10^{-13} \text{ STP J}^{-1}$ determined directly above the vent⁶. Clearly this calculation is limited by our inability to determine accurately the sulphide precipitation temperature, but the similarity between $\delta^{34}\text{S}_{\text{CDT}}$ values of these sulphides (1.4–2.2‰) and those of the vent fluid (1–1.5‰; ref. 24), and the location of all but one of the sulphides from the interior portion of the chimneys, suggest that assuming a precipitation temperature of 350 $^\circ\text{C}$ is not unreasonable.

An alternative approach is to assume a helium/heat ratio and use equation (1) or (2) to calculate a mean trapping temperature for the fluids. We have added to Fig. 1b a scale of temperatures corresponding to our estimated helium/heat ratio. As expected, sulphides from the black smokers yield higher temperatures than the sulphate from the white smoker. Sulphide ALV 980 R20 shows a significantly higher $^{40}\text{Ar}/^4\text{He}$ ratio than the others and a correspondingly lower trapping temperature. The presence of sea water in later secondary inclusions could account for this observation. Mixing of sea water and hydrothermal fluids in variable proportions has been proposed to account for small-scale variations in sulphur isotope compositions of seafloor hydrothermal sulphides²⁵.

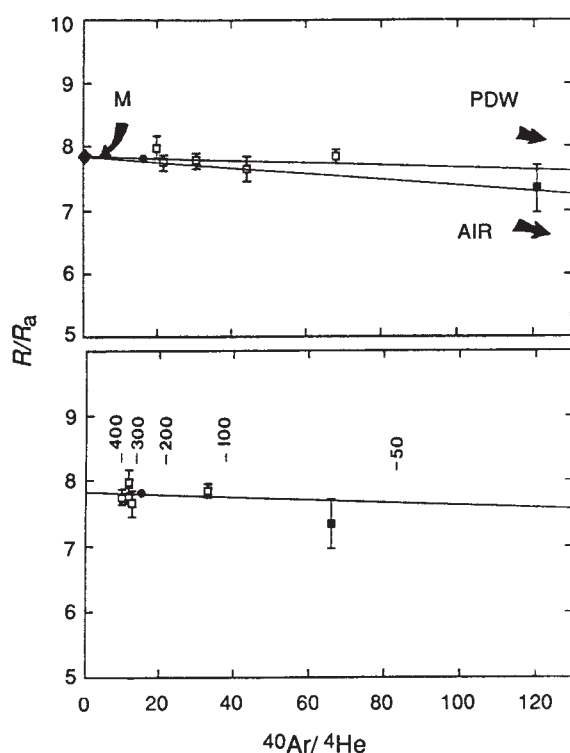


FIG. 1 Plot of R/R_a (the sample $^3\text{He}/^4\text{He}$ ratio normalized to the atmospheric value of 1.39×10^{-6}) against $^{40}\text{Ar}/^4\text{He}$. □, Black smoker sulphides; ■, white smoker sulphate; ●, contemporary vent fluid composition. a, Raw data. Mixing lines between solid diamond M (the composition of mantle gas) and Pacific deep water (PDW) and air are shown. b, Data corrected for the presence of trapped atmospheric ^{40}Ar using the Kr/Ar ratios. The numbers above the M–PDW mixing line illustrate how the $^{40}\text{Ar}/^4\text{He}$ ratios depend on the mean trapping temperature ($^\circ\text{C}$) of the fluids.

The above discussion assumes that $F(\text{Kr})$ for PDW can be used to apply a correction for trapped modern air and hence deduce $^{40}\text{Ar}/^{4}\text{He}$ ratios for the vent fluids. Elemental fractionation may occur during boiling, because solubility increases with atomic mass, and could render the correction invalid. At high temperatures, however, solubilities of the different noble gases converge²⁶, generating negligible fractionation²⁷. Furthermore, vent fluids from the southern Juan de Fuca ridge have relative abundance patterns identical to those of air-saturated sea water at 2 °C, even though boiling has led to gas loss²⁸. In contrast, vent fluids at 21° N have Ne concentrations comparable to those of Pacific deep water⁶ and thus reveal no depletion that would result from phase separation.

Detailed interpretations of He isotope and helium/heat ratios in vent fluids at 21° N have been adequately treated elsewhere^{6,8}. Here, our purpose is rather to demonstrate that measurements currently made by sampling deep ocean water can be made on trapped fluids in sulphides from the sea floor. This work could be extended to fluids in a wide range of hydrothermal deposits as a way of studying helium/heat regimes and mantle $^3\text{He}/^4\text{He}$ ratios past and present. Future work on transparent minerals will allow noble gas extractions to be combined with microthermometric measurements of trapping temperatures in individual inclusions. With opaque minerals, it will be necessary to determine precipitation temperatures from microthermometric measurements of fluids trapped by coexisting transparent minerals and associate this with the average $^{40}\text{Ar}/^4\text{He}$ ratio. Equation (2) does not rely explicitly on the measurement of ^3He and could be applied, with simultaneous laser ablation studies of H_2O and the noble gases²⁹, to the analysis of individual fluid inclusions, where ^3He is below the limits of detection. But like equation (1), this assumes a two-component mixture and

would be invalidated by the presence of radiogenic ^4He and ^{40}Ar , and it would be necessary to establish the presence of mantle $^3\text{He}/^4\text{He}$ ratios by bulk measurements. □

Received 20 March; accepted 5 May 1992.

1. Canadian-American Seamount Expedition *Nature* **313**, 212–214 (1985).
2. Lupton, J. E., Baker, E. T. & Massoth, G. J. *Nature* **337**, 161–164 (1989).
3. O'Nions, R. K. & Oxburgh, E. R. *Nature* **306**, 429–431 (1983).
4. Sclater, J. G. *et al.* *J. geophys. Res.* **86**, 11535–11552 (1981).
5. Jenkins, W. J. *et al.* *Nature* **272**, 156–158 (1978).
6. Lupton, J. E. *et al.* *Earth planet. Sci. Lett.* **50**, 115–127 (1980).
7. Lupton, J. E. *Eos* **64**, 723 (1983).
8. Weihan, J. A. & Craig, H. in *Hydrothermal Processes at Seafloor Spreading Centers* (eds Rona, P. A., Bostrom, K., Laubier, L. & Smith, K. L.) (Plenum, New York, 1983).
9. Craig, H., Weihan, J. A. & Hilton, D. R. *Eos* **68**, 1553 (1987).
10. Rosenberg, N. D. *et al.* *Nature* **334**, 604–607 (1988).
11. Baker, E. T. & Lupton, J. E. *Nature* **346**, 556–558 (1990).
12. Simmons *et al.* *Nature* **329**, 429–432 (1987).
13. RISE Project group *Science* **207**, 1421–1449 (1980).
14. Allegre, C. J., Staudacher, Th. & Sarda, Ph. *Earth planet. Sci. Lett.* **81**, 127–150 (1986/87).
15. Staudacher, Th. *et al.* *Earth planet. Sci. Lett.* **96**, 119–133 (1989).
16. Stuart, F. M. & Turner, G. *Chem. Geol.* (in the press).
17. Clarke, W. B., Beg, M. A. & Craig, H. *Earth planet. Sci. Lett.* **6**, 212–220 (1969).
18. Bieri, R. H. & Koide, M. *J. geophys. Res.* **77**, 1667–1676 (1972).
19. Bohlke, J. K. & Irwin, J. J. *Geochim. cosmochim. Acta* **56**, 187–203 (1992).
20. Smith, S. P. & Kennedy, B. M. *Geochim. cosmochim. Acta* **47**, 503–515 (1983).
21. Kelley, D. S. & Delaney, J. R. *Earth planet. Sci. Lett.* **83**, 53–66 (1987).
22. Cowan, J. & Cann, J. *Nature* **333**, 259–261 (1988).
23. Ozima, M. & Zashu, S. *Earth planet. Sci. Lett.* **62**, 24–40 (1983).
24. Woodruff, L. & Shanks, W. C. III *J. geophys. Res.* **93**, 4562–4572 (1988).
25. Crowe *et al.* *Chem. Geol.* (in the press).
26. Potter, R. & Clyne, M. A. *J. Solution Chem.* **7**, 837–844 (1976).
27. Mazar, E. & Truesdell, A. H. *Geothermics* **13**, 91–102 (1984).
28. Kennedy, B. M. *Geochim. cosmochim. Acta* **52**, 1929–1935 (1988).
29. Burnard, P. *et al.* *Proc. 7th Conf. Water-Rock Interaction 7* (in the press).

ACKNOWLEDGEMENTS. We thank the pilots and scientists aboard the 1979 and 1981 ALVIN expeditions and P. Duller for the provision of the samples. We thank D. J. Blagburn for technical assistance, P. Schofield for XRD analyses and J. Maynard for sulphur isotope measurements. Financial support from NERC and NSF is gratefully acknowledged.

Crustal control of ridge segmentation inferred from observations of the Reykjanes Ridge

R. E. Bell & W. R. Buck

Lamont-Doherty Geological Observatory of Columbia University,
Palisades, New York 10964, USA

LARGE-AMPLITUDE variations in topography and inferred crustal thickness along the axes of mid-ocean ridges, often referred to as segmentation¹, are mainly observed at slow-spreading ridges^{2–4}. This observation has led to the suggestion that mantle processes give rise to segmentation only when spreading rates are low^{5,6}. Here we make the alternative proposal that the development of segmentation is controlled by the temperature of the crust: segmentation cannot develop when the lower crust is hot enough to undergo rapid ductile flow. Thermal models predict that thick crust at a slow-spreading ridge may be as hot as normal-thickness crust along fast-spreading ridges; we accordingly test our hypothesis at a slow-spreading ridge characterized by thick crust—the Reykjanes Ridge. Topography and gravity data along the Reykjanes Ridge axis indeed show an absence of segmentation, suggesting that the thermal state of the crust, rather than any mantle process, controls the development of this structure.

The structure of mid-ocean ridges has long been inferred from gravity and bathymetry data^{7,8}. Recently, long sections of both fast- and slow-spreading ridges have been surveyed with multi-beam bathymetry and high-resolution gravimeters^{1,2,3,9,10}. The results show that along-axis variations in gravity and topography at slow ridges are generally much greater than at fast-spreading ridges. The distinctive gravity and topography signature of slow-

spreading ridges can be explained by along-axis crustal thickness variations of up to three kilometres¹. In contrast, large crustal thickness variations are not required to explain the small along-axis variations in gravity and topography documented at fast-spreading ridges⁹.

Very different processes must act at slow- and fast-spreading ridges to explain the contrasting axial crustal structures inferred from topography and gravity data. Parmentier and Phipps-Morgan⁵ suggest that the mode of mantle flow is dependent on spreading rate: fast-spreading ridges are dominated by sheet-like upwelling and slow-spreading ridges are dominated by diapiric upwelling (Fig. 1a). Sheet-like upwelling at fast ridges will supply a constant amount of melt along axis resulting in a constant crustal thickness. Diapiric upwelling along slow-spreading ridges should cause greater melting and thicker axial crust at the centres of upwelling. A second possibility is that diapiric mantle upwelling is ubiquitous, but at some ridges crustal flow evens out differences in crustal thickness. At a fast-spreading ridge, the hot crust may flow too rapidly to maintain crustal thickness variations, whereas at a slow-spreading ridge, the cold crust locks in variations in thickness resulting from diapiric mantle flow (Fig. 1b).

At a slow-spreading ridge with thick and presumably hot axial crust, the model based on differences in mantle flow would predict large along-axis variations in crustal thickness, whereas the model based on differences in crustal flow would predict little variation. Here we use gravity and topography data from the hot, slow-spreading Reykjanes Ridge to test for along-axis crustal thickness variations.

A mantle Bouguer anomaly is constructed from a free-air gravity anomaly by removing the gravitational effect of the crust-water interface and the effect of a constant-thickness crust⁹. The resulting anomalies reflect deviations from the assumed density structure of the crust or mantle. Along 325 km of the Mid-Atlantic Ridge, south of the Atlantis Fracture Zone,

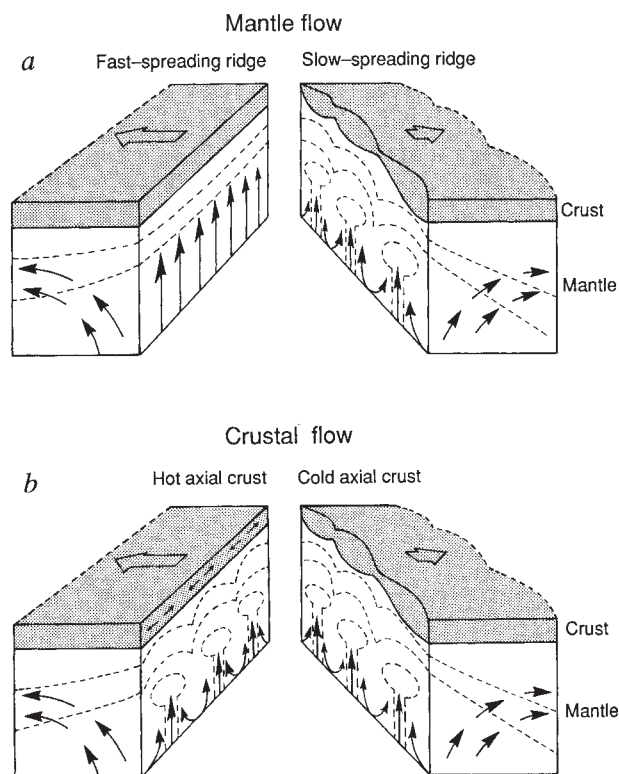


FIG. 1 Schematic illustration of two mechanisms for producing along-axis segmentation at slow-spreading ridges. *a*, Mantle flow model. In this model the segmentation is the result of diapiric mantle upwelling beneath slow-spreading ridges (right) represented by heavy arrows resulting in regions of focused upward flow causing thicker crust above upwelling centres. Beneath fast-spreading ridges (left), the upward flow of the mantle is uniform or sheet-like resulting in constant crustal thickness along the axis. The crust is simply a passive recorder of the changing patterns of mantle upwelling. The figure is modified from ref. 6. *b*, Crustal flow model. In this model the upwelling mantle is diapiric beneath all ridges. The variations in crustal thickness caused by this diapirism are only preserved when the crustal flow is sluggish (that is when the crust is cold, as for the case at the right). When the axial crust is hot, it can flow fast enough to even out any significant crustal thickness variations, as shown at the left. The arrows within the crustal section of the hot axial crust illustration indicate this flow of material away from the upwelling centres.

Lin *et al.*¹ calculated a mantle Bouguer anomaly with a full two-dimensional data set. We have used the axial bathymetry and free-air gravity data from this area to demonstrate the validity of a calculation based on a single axial profile. The segmentation recognized in the two-dimensional data is clearly visible in the single along-axis profile (Fig. 2). Some of the difference between the simple axial calculation and the two-dimensional treatment probably arises from variations in the rift flank topography. The smoothness of the axial topography along the Reykjanes Ridge should minimize the contribution of this potential error source.

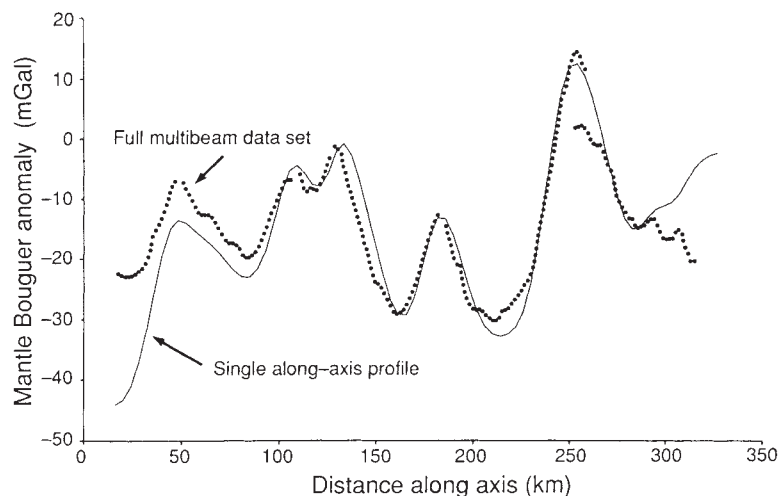
The Reykjanes Ridge is the portion of the Mid-Atlantic Ridge extending over 1,200 km from Iceland south to the Charlie Gibbs Fracture Zone (Fig. 3). The Reykjanes Ridge, spreading at a half rate of 1 cm yr^{-1} (refs 7, 11), lacks the distinct axial valley characteristic of most slow-spreading ridges. The axial high along the Reykjanes Ridge resembles the topography of the fast-spreading East Pacific Rise. The crust along the ridge varies

from 8 to 10 km thickness^{12,13} in contrast to normal oceanic crustal thicknesses of 5–6 km (ref. 14). The Reykjanes Ridge has an average depth of 1,000 m, whereas the average depth of the Mid-Atlantic Ridge south of the Atlantis Fracture Zone is $\sim 3,000 \text{ m}$ (refs 1, 10). The isotope signature of basalts along the axis of the Reykjanes Ridge indicates that the thick crust and shallow water depth are associated with the Iceland hotspot¹⁵.

Our geophysical data were collected during a RV *Vema* cruise along the Reykjanes Ridge (Fig. 3). For 83 crossings made during the entire cruise, the r.m.s. errors were 4.2 mGal with a mean of 0.9 mGal for the gravity and 58.6 m with a mean of 2.5 m for the topographic data. Analysis of the profiles evaluated here results in r.m.s. errors of 3.6 mGal and 64.3 m for seven crossovers.

Along the Reykjanes Ridge, we calculated mantle Bouguer anomalies from both a single along-ridge profile centred along the axial magnetic anomaly and the long lines perpendicular to

FIG. 2 Mantle Bouguer anomaly along the ridge axis south of the Atlantis Fracture Zone. The dotted line is the axial profile extracted from a full two-dimensional study⁵. The solid line is the mantle Bouguer anomaly calculated from a single along-axis profile resampled from the Lin *et al.*¹ data.



the ridge using a crustal density of $2,700 \text{ kg m}^{-3}$ and a 9-km-thick crust. The r.m.s. difference between the axial mantle Bouguer anomaly and the ridge perpendicular calculations is 5.2 mGal for the seven crossovers (Fig. 4a). The striking features of the Reykjanes mantle Bouguer anomaly are the small-amplitude (5–10 mGal) short-wavelength (10–25 km) anomalies superimposed a regional gradient, and the general absence of high-amplitude (10–60 mGal) long-wavelength (35–75 km) variations observed along other slow-spreading ridges.

These short-wavelength anomalies either arise from density variations in the upper crust or are artefacts of the one-dimensional assumption inherent in the calculation. A 5% density variation in the upper crust or a 200-m topographic feature ignored in the one-dimensional calculation could produce this signal. A Moho source for these anomalies is unlikely because of the amplitude and wavelength of crustal thickness variations required. Forward modelling demonstrates that the crustal thickness variations required to explain these anomalies are double the thickness variations proposed along other slow-spreading ridges and much shorter wavelength.

The uniqueness of the mantle Bouguer anomaly along the Reykjanes Ridge is best illustrated by comparisons with other mid-ocean ridges (Fig. 4b). The Reykjanes Ridge lacks the large anomalies observed at slow-spreading ridges in the North and South Atlantic. We have upward continued the mantle Bouguer anomaly to a distance of 3,000 m from the sea floor to simulate an anomaly along a typical, deeper mid-ocean ridge. This upward continued Reykjanes Ridge anomaly retains a higher-frequency signal than the fast-spreading East Pacific Rise. The Reykjanes Ridge also lacks large (700–1,500 m) variations in axial depth observed along other slow-spreading ridges. Thus neither the gravity nor the topography indicates along-axis variations in the crustal thickness at the Reykjanes Ridge.

Other characteristics of this slow-spreading ridge are its anomalously thick crust and associated higher crustal temperatures. Crust formed at fast-spreading ridges is also hotter than crust along typical slow-spreading ridges. These observations imply that the temperature of the crust controls whether large-amplitude segmentation develops along a ridge. The segmentation of crustal thickness along ridges may be a

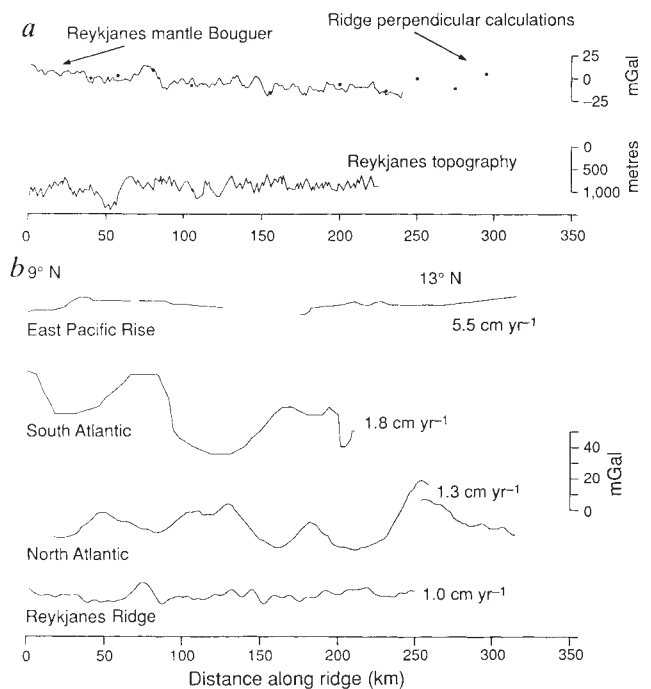


FIG. 4 a, Mantle Bouguer anomaly along the Reykjanes Ridge. The upper profile is the result of the axial profile calculations and the dots result from the calculations of the axis perpendicular lines. The axial topography profile is below the mantle Bouguer anomaly. b, Mantle Bouguer anomalies along spreading ridges with a wide variety of spreading rates. The absolute values of the profiles were offset to facilitate the comparison. The top profiles are from two locations along the East Pacific Rise². The left-hand values are from 9° N and the right-hand samples are from 13° N. The second profile is the mantle Bouguer anomaly calculated by Kuo and Forsyth⁹ from the South Atlantic. The mantle Bouguer anomaly calculated by Lin *et al.*¹ for the region in the North Atlantic south of the Atlantis Fracture Zone is next. The Reykjanes Ridge anomaly, with the long-wavelength trend removed, and upward continued to simulate an average axial depth of 3,000 m, is illustrated at the bottom.

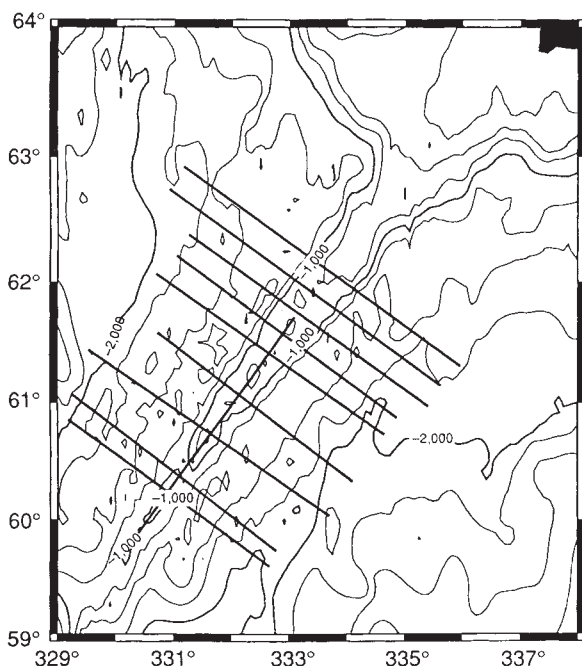


FIG. 3 Bathymetry of the Reykjanes Ridge, south of Iceland, contoured at 250-m intervals. Iceland is in the northeast corner. The tracklines from the Vema 23 cruise⁷ are superimposed on the bathymetry.

consequence of mantle diapirism or may be related to lithospheric processes. Whatever mechanism causes segmentation at slow-spreading ridges, it cannot produce segmentation wherever crustal temperatures are evaluated.

The strength and viscosity of crustal rocks is a strong function of temperature. Because the crust of the Reykjanes Ridge is much thicker than typical oceanic crust, the base of the crust should be much hotter than at other slow-spreading ridges. A simple analytical model can be used to illustrate the effect of plate velocity and crustal thickness on the lower crustal temperature. In this one-dimensional model we assume a constant upward velocity equal to the plate velocity. The boundary conditions applied to the heat-flow equation are that the upper surface is at 0 °C and the temperature at infinite depth is 1,300 °C. The thermal diffusivity is taken to be $10^{-6} \text{ m}^2 \text{ s}^{-1}$. For a plate velocity of 1 cm yr^{-1} a crustal temperature greater than 1,150 °C will only be achieved for thicknesses greater than 6.8 km. At a faster rate of 5 cm yr^{-1} this temperature is reached at 1.36 km depth. Many effects could alter these numbers, including hydrothermal circulation and advection of magma¹⁶.

Thick crust at slow-spreading ridges or normal crust at fast-spreading ridges may be hot enough partially to melt, and a steady-state magma chamber may be present. Even if a steady-state magma chamber is absent, the lower crust may be weak enough to flow rapidly. To estimate the conditions needed for flow to remove topographic variations, we approximate the flow of a crustal layer as a channel flow^{17,18}. Because crust cools as it moves away from a spreading centre, there is limited time for

crustal flow to operate. A continuous magma chamber as thin as 3 cm filled with basaltic melt having a viscosity of 10 Pa s could remove 50-km-wavelength topographic variations within 10,000 years. Alternatively, if 3 km of lower crust had a viscosity of 10^{16} Pa s then topography would decay at the same rate.

It has been suggested that crustal rheology determines the presence or absence of axial valleys^{19,20}. Axial valleys are

prominent along most slow-spreading ridges, but are absent along most fast-spreading ridges. The presence of either a long-lived magma chamber or very weak crust is thought to preclude the formation of a prominent axial valley. Our results suggest that crustal flow is as important in crustal segmentation as it is in determining the form of across-ridge topography. □

Received 3 March; accepted 23 April 1992.

- Lin, J., Purdy, G. M., Schouten, H., Sempere, J. C. & Zervas, C. *Nature* **344**, 627–632 (1990).
- Madsen, J. A., Detrick, R. S., Mutter, J. C., Buhl, R. & Orcutt, J. A. *J. geophys. Res.* **95**, 4967–4987 (1990).
- Morris, E. & Detrick, R. S. *J. geophys. Res.* **96**, 4355–4366 (1991).
- Blackman, D. K. & Forsyth, D. W. *J. geophys. Res.* **96**, 11741–11758 (1991).
- Parmentier, E. M. & Phipps-Morgan, J. *Nature* **348**, 325–328 (1990).
- Lin, J. & Phipps-Morgan, J. *Geophys. Res. Lett.* **19**, 13–16 (1992).
- Talwani, M., Windisch, C. C. & Langseth, M. G. *J. geophys. Res.* **76**, 473–517 (1971).
- Woodside, J. M. *Can. J. Earth Sci.* **9**, 942–959 (1972).
- Kuo, B. & Forsyth, D. W. *Mar. geophys. Res.* **10**, 205–232 (1988).
- Sempere, J.-C., Purdy, G. M. & Schouten, H. *Nature* **344**, 427–431 (1990).
- Vogt, Peter R. in *The Geology of North America*, Vol. M (eds Vogt, P. & Tucholke, B.) 189–204 (Geological Society of America, Boulder, 1986).

- Bunch, A. & Kennett, B. L. N. *Geophys. J. R. astr. Soc.* **61**, 141–166 (1980).
- RRISP *J. Geophys.* **47**, 228–238 (1980).
- Purdy, G. M. and Ewing, J. in *The Geology of North America*, Vol. M (eds Vogt, P. & Tucholke, B.) 313–330 (Geological Society of America, Boulder, 1986).
- Schilling, J. G. *The Geology of North America*, Vol. M (eds Vogt, P. & Tucholke, B.) 137–156 (Geological Society of America, Boulder, 1986).
- Chen, Y. & Morgan, W. J. *J. geophys. Res.* **95**, 9275–9282 (1989).
- Turcotte, D. L. & Schubert, G. *Geodynamics: Applications of Continuum Physics to Geological Problems* (Wiley, New York, 1982).
- Bird, P. *J. geophys. Res.* **96**, 10275–10286 (1991).
- MacDonald, K. C., Scheirer, D. S. & Carbotte, S. M. *Science* **253**, 986–994 (1991).
- Wilson, D. S., Clague, D. A., Sleep, N. H. & Morton, J. L. *J. geophys. Res.* **93**, 11974–11984 (1988).

ACKNOWLEDGEMENTS. We thank J. Cochran, J. Mutter, J. Lin, M. Spiegelman and P. Flemings for their comments. This work was supported by the NSF and the Office of Naval Research.

Desert ants on a thermal tightrope

R. Wehner*, A. C. Marsh† & S. Wehner*

*Department of Zoology, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland

†Department of Zoology, University of Namibia, PB 13301, Windhoek, Namibia

MANY animals restrict their foraging activities to certain times of the day or night, but the Saharan silver ant *Cataglyphis bombycina* is exceptional in that all foragers leave their underground nest in an explosive outburst confined to a few minutes per day during the hottest midday period. The foraging activity of this 'thermophilic' ant is compressed into a small thermal window by predatory pressure on the one hand and heat stress on the other.

On a summer day in the central Sahara, the silver ant, *C. bombycina*, is the only arthropod that forages in the full midday sun, even when surface temperatures exceed 60 °C (refs 1, 2). As a scavenger, it searches for other arthropods that were active during the night and early morning but did not retreat quickly enough into their daytime shelters, and consequently succumbed to the heat and desiccation stress caused by the rising sun. In contrast to *C. bombycina*, other species of desert ants stop foraging and retreat to their underground burrows at surface temperatures of 35–45 °C (refs 3–8). The silver ants, however,

start their foraging activity only then and continue to forage until their body temperature reaches the critical thermal maximum of 53.6 ± 0.8 °C (for definition and methods, see legend to Fig. 1a). This is the highest critical thermal maximum value recorded so far for any terrestrial animal^{9,10}.

Furthermore, although other *Cataglyphis* species inhabiting more mesic habitats, for example *C. bicolor*, are active during the entire course of the day (Fig. 1a)¹¹, *C. bombycina* confines its foraging activity to an extremely narrow time range. During this period, a colony's forager force, consisting of a few hundred individuals, leaves its nest in a dramatic 'outburst' lasting for only a few minutes ($t_1 = 3.6 \pm 0.6$ min; Fig. 1b). This outburst occurs when the temperature measured at ant height (4 mm above the surface) has risen to 46.5 ± 2.1 °C (Table 1). Owing to the ant's small body size (9.72 ± 2.63 mg) and resultant low thermal inertia, its body temperature will be very close to the temperature measured at ant height^{12,13}. (Vertical temperature gradients above ground were measured at 15-min intervals using an array of copper-constantan thermocouples mounted at 0, 2, 4, 6, 10, 20, 100 and 200 mm above the sand surface and connected to a data logger.)

As a result, the foraging activity of *C. bombycina* is compressed into a small thermal window, with a maximum width of 7 °C (46.5 – 53.6 °C; Fig. 2A). This thermal window is so close to the ant's upper lethal temperature that the animals must spend at least 30 per cent, and sometimes up to 75 per cent, of their foraging time in thermal refuges (Fig. 2A) by pausing on the tops of stalks of dry vegetation where they encounter much

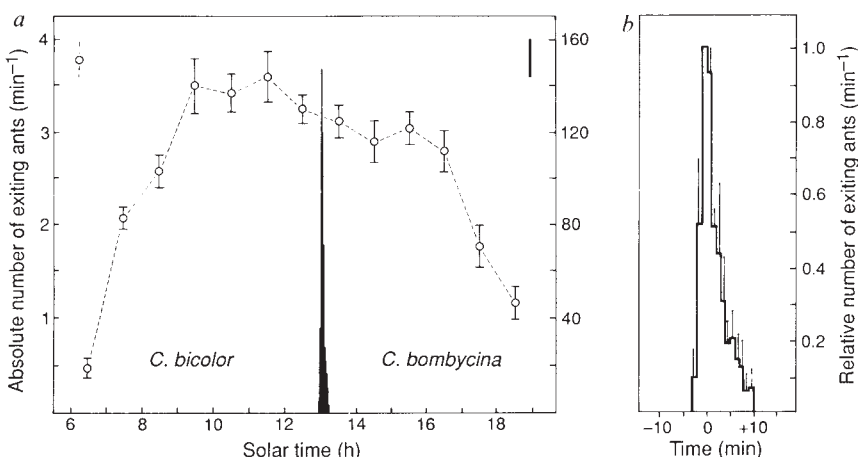


FIG. 1 a, Diel foraging activity patterns of *Cataglyphis bicolor* (left ordinate) and *C. bombycina* (right ordinate) recorded in the shrub desert of the North African sahel and the Saharan sand-dune desert, respectively. Foraging activity is expressed as numbers of ants leaving the nest per minute. In *C. bicolor*, means $\pm$ s.d. are given for 12 nests. In *C. bombycina*, the figure depicts the foraging 'outburst' of one individual nest. b, Time course of the foraging outburst of *C. bombycina* as computed from the data of four outbursts recorded on four consecutive days at one nest (mean $\pm$ s.d.).

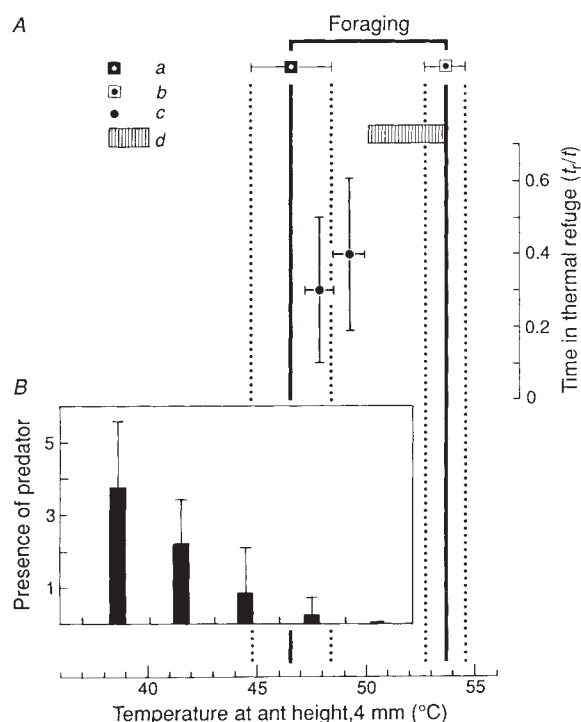


FIG. 2 Narrow thermal window into which the foraging activity of *C. bombycina* is compressed by heat stress (upper limit) and predation pressure (lower limit). The latter is caused by the ant-eating lizard, *Acanthodactylus dumerili* (B). Symbols in A: a, outburst (start of foraging activity); b, critical thermal maximum (CT_{max}). CT_{max} values were measured in the laboratory¹⁰ by exposing individual ants to a rising temperature gradient of 1°C min^{-1} . Within this gradient, CT_{max} was defined as that temperature at which the animal curled its metasoma tightly underneath the thorax and started to walk in a jerky uncoordinated way. Finally, *C. bombycina* died when the rising temperatures reached $55.2 \pm 1.1^\circ\text{C}$; c, fraction of total foraging time spent by the ants in thermal refuges (right ordinate: t_r , time in thermal refuge; t , total foraging time); d, range of temperatures at which foraging ants were observed to succumb to heat stress (when thermal refuge sites were not available along the ant's foraging path). B, Number of lizards seen outside their burrows along a 250-m transect. Means $\pm$ s.d. are given for all data.

lower temperatures and can off-load excess body heat. These physical constraints imposed on the animals by the need of thermoregulatory behaviour markedly limit the amount of extranidal time available for foraging. If the ants are unable to find and exploit such refuge sites during their foraging excursions they die while moving about on the desert floor. Consequently, while foraging and homing, *C. bombycina* is totally dependent on the opportunity to perform thermal respite behaviour, namely to lower its body temperature by exploiting the microclimatic mosaic that occurs even within its flat and barren sand-floor habitat.

In trying to understand this unusual foraging behaviour of the Saharan silver ants, two questions spring to mind. (1) What triggers the outburst and thus starts the ants' foraging activity? (2) Why do the ants restrict their foraging activity to such high temperatures? An answer to the first question derives from the observation that, before the outburst, individual ants regularly appear above ground where they apparently monitor the ambient temperature. When the threshold temperature (at ant height) is reached, these individuals trigger the outburst by stimulating the forager force that is waiting in the uppermost chamber of the nest. In this respect, it is important to note that *C. bombycina* is the only *Cataglyphis* species in which mandibular gland secretions have been found to elicit strong stimulation of nestmates¹⁴.

TABLE 1 Correlation between the occurrence of the foraging outbursts of *Cataglyphis bombycina* and temperatures recorded at various heights above ground

Test period	Ambient temperature ($^\circ\text{C}$), daily maximum	Outburst temperature ($^\circ\text{C}$) at			n
		0 mm	4 mm	200 mm	
1	48.7 ± 1.1	50.0 ± 2.5	46.8 ± 1.9	40.8 ± 2.7	19
2	39.6 ± 1.8	57.1 ± 2.0	46.0 ± 2.3	35.1 ± 2.4	12

Test periods 1 and 2 refer to hot and (relatively) cool days, respectively (see second column; ambient temperature is measured conventionally at 1.0 m above ground, in the shade). Irrespective of the different temperature regimes in the two test periods, the ants' outbursts occur at the same ant-height (at 4 mm) temperature ($P > 0.25$), but there is no correlation between the outbursts and either surface (at 0 mm) temperature or temperature at 200 mm: 0-mm and 200-mm temperatures differ at the time of the outbursts between test periods 1 and 2 (in either case, $P < 0.001$). n, Number of outbursts.

The second question focuses on the ultimate reason for the ants' hazardous behaviour. Why does this particular species of ant run the high risk of getting overheated by foraging just below lethal temperatures? The answer lies in the predation pressure exerted on the ants by a desert lizard, *Acanthodactylus dumerili*¹⁵⁻¹⁸. During the hottest times of a summer day, this lizard concentrates on *C. bombycina* as its only potential prey, because then the silver ants are the only arthropods foraging above ground. The lizard, which frequently has its burrows close to the *C. bombycina* colonies, retreats underground at about the temperature at which the ants' foraging activity begins (Fig. 2B) (for heat tolerances of *Acanthodactylus* spp. and other lizards, see refs 19-22). If ants are experimentally released on the desert floor before the natural outbursts of the colonies occur, they fall victim to their predators often within less than five minutes.

The small overlap between the lizard's and the ant's natural foraging times (Fig. 2A, B) indicates that the ant-eating lizard has ultimately shaped the ant's thermal, and thus temporal, foraging behaviour. In conclusion, *C. bombycina* is forced to exploit an extremely narrow thermal window in which the upper and lower limits are set by heat stress and predatory pressure, respectively. □

Received 16 March; accepted 8 May 1992.

1. Larmuth, J. *Sahara Desert* (ed. Cloudsley-Thompson, J. L.) 57-66 (Pergamon, Oxford, 1984).
2. Coenen-Stass, D. *Monit. Zool. Ital. (NS) Monogr.* **4**, 425-434 (1989).
3. Clark, W. H. & Comanor, P. L. *Amer. Midl. Naturalist* **90**, 467-474 (1973).
4. Rogers, L. E. *Environ. Entomol.* **3**, 420-424 (1974).
5. Schumacher, A. & Whitford, W. G. *Insectes Sociaux* **21**, 317-330 (1974).
6. Briesse, D. T. & MacAuley, B. J. *Aust. J. Ecol.* **5**, 121-134 (1980).
7. Morton, S. R. *Evolution of the Flora and Fauna of Arid Australia* (ed. Barker, W. R. & Greenslade, P. J. M.) 257-262 (Peacock, Adelaide, 1982).
8. Marsh, A. C. *J. arid Environ.* **14**, 61-73 (1988).
9. Curtis, B. A. *J. Insect Physiol.* **31**, 463-466 (1985).
10. Marsh, A. C. *Physiol. Zool.* **58**, 626-636 (1985).
11. Wehner, R., Harkness, R. D. & Schmid-Hempel, P. *Foraging Strategies of Individually Searching Ants, Cataglyphis bicolor (Hymenoptera: Formicidae)* (Fischer, Stuttgart, 1983).
12. Bakken, G. S. *J. theor. Biol.* **60**, 337-384 (1976).
13. Casey, T. M. *Insect Thermoregulation* (ed. Heinrich, B.) 80-114 (Wiley, New York, 1981).
14. Hefetz, A. & Lloyd, H. A. *Z. Naturforsch. C* **40**, 665-666 (1985).
15. Salvador, A. *Bonner Zool. Beitr.* **16**, 1-167 (1982).
16. Bons, J. *Bull. Soc. Sci. Nat. Maroc* **38**, 167-182 (1959).
17. Blanc, C. P. & Sciortino, P. *Bull. Soc. Sci. Nat. Tunisie* **13**, 25-38 (1978).
18. Noura, S. *Bull. Soc. Zool. France* **108**, 477-483 (1983).
19. Busack, S. D. *Copeia* **76**, 826-830 (1976).
20. Pough, R. H. & Busack, S. D. *J. therm. Biol.* **3**, 203-205 (1978).
21. Bradshaw, S. D. *Ecophysiology of Desert Reptiles* (Academic, Sydney, 1986).
22. Pianka, E. R. *Ecology and Natural History of Desert Lizards* (Princeton Univ. Press, New Jersey, 1986).

ACKNOWLEDGEMENTS. We thank V. Domanico for her competence and help in the laboratory. This work was supported by a grant of the Swiss National Science Foundation.

Involvement of the epidermal growth factor receptor in the invasion of cultured mammalian cells by *Salmonella typhimurium*

Jorge E. Galán, John Pace & Michael J. Hayman

Department of Microbiology, School of Medicine, State University of New York at Stony Brook, Stony Brook, New York 11794, USA

SALMONELLA infection continues to be a major world-wide health problem¹. One essential pathogenic feature common to all *Salmonella* is their ability to penetrate the cells of the intestinal epithelium which are normally non-phagocytic². The internalization of *Salmonella* into mammalian cells is thought to be a receptor-mediated phenomenon and the invasion of cultured epithelial cells depends on several *Salmonella* genes, but nothing is known about the host determinants participating in this interaction³⁻⁶. Protein tyrosine phosphorylation follows stimulation of many cell-surface receptors to initiate signal transduction pathways that stimulate cellular responses⁷. We report here that invasion of cultured Henle-407 cells by *Salmonella typhimurium* induces the tyrosine phosphorylation of the epidermal growth factor (EGF) receptor. In contrast, an isogenic strain of *S. typhimurium* that is defective in invasion owing to a mutation in the *invA* gene is unable to induce such phosphorylation. Addition of EGF to cultured Henle-407 cells allowed the internalization of the invasion-defective *S. typhimurium invA* mutant although it did not cause the internalization of an adherent, but non-invasive, strain of *Escherichia coli*. This result indicates that stimulation of the EGF receptor is involved in the invasion of cultured Henle-407 cells by *S. typhimurium*.

We investigated the tyrosine phosphorylation of proteins in cultured Henle-407 human epithelial cells after challenge with

a wild-type virulent strain of *S. typhimurium*. *S. typhimurium* infection increased the tyrosine phosphorylation of several proteins, including one prominent species of relative molecular mass 170,000 (170K) (Fig. 1, lane C). Cultured cells infected with the isogenic mutant strain of *S. typhimurium* SB148 failed to increase the tyrosine phosphorylation of this protein (Fig. 1, lane E). This strain carries an insertion mutation in the gene *invA* which renders it deficient in its ability to invade cultured epithelial cells, although it can still fully attach to these cells^{6,8}. In addition, this mutant is unable to cause host cell cytoskeletal rearrangements and to induce changes in the concentrations of host free intracellular calcium ($[Ca^{2+}]_i$) (ref. 8).

The EGF receptor is one of a family of receptors with tyrosine kinase activity⁷. On binding EGF, this receptor undergoes rapid tyrosine autophosphorylation on several residues and activation of its tyrosine kinase. This activity is essential for several responses of the cell to the growth factor, including membrane ruffling⁹, reorganization of actin microfilaments¹⁰ and a rapid increase in $[Ca^{2+}]_i$ ¹¹, which are cellular events also observed in cultured cells infected with *Salmonella*³. We examined Henle-407 cells for the presence of a functional EGF receptor. Addition of EGF to the cells caused rapid tyrosine phosphorylation of a variety of proteins, including a 170K polypeptide that could be immunoprecipitated by a monoclonal antibody directed against the EGF receptor (data not shown; Fig. 2, lanes A and B). Furthermore, addition of EGF induced the rapid formation of membrane ruffles on Henle-407 cells (data not shown), a response that is accompanied by cytoskeleton reorganization and Ca^{2+} mobilization⁹⁻¹¹. We therefore tested whether the 170K polypeptide, which is tyrosine phosphorylated upon challenge with *S. typhimurium*, was the EGF receptor. As shown in Fig. 2 lanes G and H, a monoclonal antibody directed to the EGF receptor efficiently immunoprecipitated the 170K tyrosine-phosphorylated polypeptide from lysates of cells challenged with *S. typhimurium*, indicating that the EGF receptor is indeed phosphorylated upon *Salmonella* infection of cultured cells.

Cells infected with the invasion-deficient *S. typhimurium invA* mutant SB148 did not show significant tyrosine phosphorylation of the 170K polypeptide (Fig. 1, lane E), even though this mutant is fully capable of attaching to cultured Henle-407 cells. This indicated that stimulation of the EGF receptor may be necessary for *Salmonella* entry into Henle-407 cells and that stimulation of the EGF receptor by EGF might rescue the invasive phenotype of the SB148 *S. typhimurium* mutant. We tested this idea by examining the effect of the addition of EGF to cells infected with the invasion-defective *S. typhimurium invA* mutant. As shown in Table 1, addition of as little as 0.03 nM EGF to Henle-407 cells successfully rescued the invasion phenotype of *S. typhimurium* SB148. To exclude the possibility that addition of EGF caused generalized endocytosis and thereby internalization of any bacteria attached to the surface of Henle-407 cells, we tested the effect of EGF on the internalization of the non-invasive strain of *E. coli* RDEC-1 (ref. 12). This pathogenic strain of *E. coli* can attach to cultured epithelial cells although it is unable to enter those cells¹². As shown in Table 1, addition of EGF did not cause internalization of *E. coli* RDEC-1, indicating that the effect is specific for *S. typhimurium*. The phenotypic rescue of *S. typhimurium* SB148 was growth-factor-specific because addition of insulin at concentrations that should stimulate the receptors of both the insulin and the insulin-like growth factor IGF-I did not trigger the internalization of this strain. The induction of membrane ruffling in the Henle-407 cells by these concentrations of insulin indicates that the receptor(s) was stimulated (data not shown).

We have presented evidence that the EGF receptor is stimulated by wild-type *S. typhimurium* infection and that this stimulation is necessary for the invasion by the organism of cultured Henle-407 cells. This constitutes another example of the subversion of the EGF receptor system for the benefit of microbial pathogens. It has been suggested that the EGF receptor is

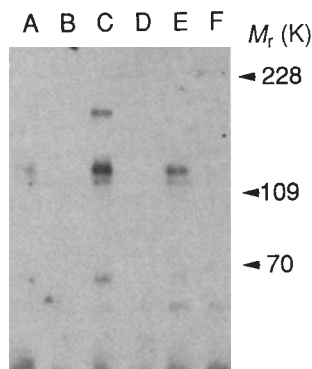


FIG. 1 Tyrosine phosphorylation of Henle-407 proteins after *S. typhimurium* infection. Samples are from uninfected cells (lanes A and B), from cells infected with wild-type *S. typhimurium* (lanes C and D) and from cells infected with the *S. typhimurium invA* mutant (lanes E and F). Cell lysates were immunoprecipitated with either a rabbit antibody against phosphotyrosine (lanes A, C and E) or with a preimmune rabbit serum (lanes B, D and F). Samples were separated on a 7% SDS-polyacrylamide gel, transferred to nitrocellulose and probed with the antiphosphotyrosine monoclonal antibody 5E2. Numbers to the right indicate the position of molecular weight standards.

METHODS. Henle-407 cells were cultured in 6-well plates to 80% confluence and, where indicated, infected with either wild-type *S. typhimurium* strain SL1344 (ref. 15) or with the *invA* isogenic mutant SB148 (refs 6, 8) at a multiplicity of infection of 100 in Hanks' balanced salt solution (HBSS). Infection was for 3 h on ice and 40 min at 37 °C. Cells were washed with HBSS containing 1 mM sodium vanadate and lysed in 0.5 ml lysing buffer (10 mM HEPES, pH 7.5, 150 mM NaCl, 10% glycerol, 1 mM $NaVO_4$, 0.6% Triton X-100, 1% aprotinin, 1 μ M PMSF) per well. Immunoprecipitation and western blotting were carried out according to ref. 17; rabbit and mouse monoclonal antiphosphotyrosine antisera have been described¹⁸.

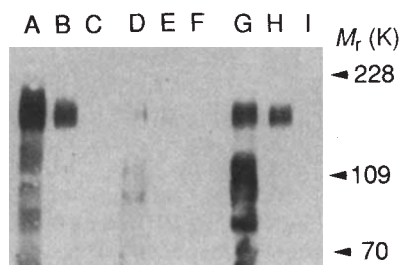


FIG. 2 Induction of tyrosine phosphorylation of the EGF receptor by *S. typhimurium*. Samples are from cells treated with EGF (lanes A, B and C), untreated and uninfected cells (lanes D, E and F) and from cells infected with wild-type *S. typhimurium* (lanes G, H and I). Cell lysates were immunoprecipitated with rabbit antibody against phosphotyrosine (lanes A, D and G), rat monoclonal against EGF receptor (lanes B, E and H) or with a preimmune rabbit serum (lanes C, F and I). Samples were separated on a 7% SDS-polyacrylamide gel, transferred to nitrocellulose membranes and probed with the antiphosphotyrosine monoclonal antibody 5E2. Numbers to the right indicate the position of molecular weight standards.

METHODS. *S. typhimurium* infection of Henle-407 cells was as described in the legend to Fig. 1, except that 100-mm tissue-culture plates were substituted for 45-mm 6-well plates. Cells were treated with 1 nM EGF for 20 min. Immunoprecipitation and western immunoblotting as described in Fig. 1 legend, except that cells were lysed in 1 ml lysis buffer. The rat monoclonal antibody against the EGF receptor is directed against the kinase domain of the chicken EGF receptor which is common to the human EGF receptor¹⁹. Rabbit and mouse monoclonal antibodies against phosphotyrosine are described in the legend to Fig. 1.

involved in the internalization of vaccinia virus¹³. In addition, a number of viruses produce EGF-like growth factors that enhance their ability to replicate in the host¹⁴. *Salmonella* has the ability to invade many cell lines (C. Ginocchio and J.E.G., unpublished results). Even though the EGF receptor is widely distributed among different tissues, it is not expressed on every cell type, so *Salmonella* may use alternative pathways to enter different eukaryotic cells, as does another invasive organism, *Yersinia sp.*¹⁵.

It is at present unclear whether *Salmonella* interacts directly with the EGF receptor or whether stimulation of the EGF receptor is a step in the internalization pathway. Tyrosine phos-

phorylation of the EGF receptor has been reported to result as a consequence of ligand binding. This raises the possibility that *Salmonella* may directly interact with the receptor. Nevertheless, pretreatment of Henle-407 cells with antibodies directed against the extracellular domain of the EGF receptor not only failed to block invasion, but rather slightly increased the number of internalized bacteria by 2–3-fold. These results can be explained by the activation of the receptor by these antibodies, presumably by causing receptor dimerization. Further investigation will establish whether a determinant(s) on the surface of *Salmonella* can bind and stimulate the EGF receptor. If this determinant is present, InvA is an unlikely candidate for this function. InvA belongs to a family of proteins involved in the translocation of specific proteins across the bacterial membrane⁸, so it is most likely to be involved in the regulation of the expression and/or localization of the effector molecule(s). Nevertheless, the results presented here will facilitate the identification of such a molecule(s). Knowledge of bacterial ligands and their receptors is central to the design of treatments and prevention of diseases caused by invasive organisms, which constitute an urgent health problem. □

Received 27 February; accepted 24 April 1992.

- Chalker, R. B. & Blaser, M. J. *Rev. infect. Dis.* **10**, 111–117 (1988).
- Takeuchi, A. *Am. J. Path.* **50**, 109–136 (1967).
- Ginocchio, C., Pace, J. & Galán, J. E. *Proc. natn. Acad. U.S.A.* (in the press).
- Finlay, B. B. *et al. Molec. Microbiol.* **2**, 757–766 (1988).
- Elsinghorst, E. A., Baron, L. S. & Kopecko, D. J. *Proc. natn. Acad. Sci. U.S.A.* **86**, 5173–5177 (1989).
- Galán, J. E. & Curtiss III, R. *Proc. natn. Acad. Sci. U.S.A.* **86**, 6383–6387 (1989).
- Ullrich, A. & Schlessinger, J. *Cell* **61**, 203–212 (1990).
- Galán, J. E., Ginocchio, C. & Costeas, P. *J. Bact.* (in the press).
- Kadowaki, T. *et al. J. biol. Chem.* **261**, 16141–16147 (1986).
- Moolenaar, W. H., Tertoolen, G. J. & de Laat, S. W. *J. biol. Chem.* **259**, 8066–8069 (1984).
- Rijken, P. J., Hage, W. J., Van Bergen En Henegouwen, P. M. P., Verkleij, A. J. & Boonstra, J. *J. Cell Sci.* **100**, 491–499 (1991).
- Cheney, C. P., Chad, P. A., Formal, S. B. & Boedecker, E. C. *Infect. Immun.* **28**, 1019–1027 (1980).
- Eppstein, D. A. *et al. Nature* **318**, 663–665 (1985).
- Todaro, G. J., Rose, T. M., Spooner, C. E., Shoyab, M. & Plowman, G. D. *Sem. Cancer Biol.* **1**, 257–263 (1990).
- Isberg, R. R. *Science* **252**, 934–938 (1991).
- Holseth, S. K. & Stocker, B. A. D. *Nature* **291**, 238–239 (1981).
- Kamps, M. P. & Sefton, B. M. *Oncogene*, **2**, 305–315 (1988).
- Lee, E. B. & Hayman, M. J. *Oncogene*, **5**, 1165–1172 (1990).
- Nair, N., Davis, R. J. & Robinson, H. L. *Molec. cell. Biol.* **12**, 2010–2016 (1992).

ACKNOWLEDGEMENTS. This work was supported by grants from the NIH, the Sinsheimer Foundation and the Searle-Chicago Community Trust. J.E.G. is a Pew scholar in the biomedical sciences.

TABLE 1 Invasion of *invA* *S. typhimurium* strain SB148 and *E. coli* RDEC-1 treated with EGF

Strain	Treatment	Invasion (%) [*]
SL1344†	None	36 ± 1
SB148‡	None	0.04 ± 0.002
SB148	EGF (16 nM)	17 ± 0.5
SB148	EGF (3 nM)	9 ± 0.3
SB148	EGF (0.3 nM)	6 ± 0.8
SB148	EGF (0.03 nM)	2 ± 0.04
RDEC-1§	None	0.003 ± 0.0002
RDEC-1	EGF (16 nM)	0.002 ± 0.0003

Henle-407 cells grown on 24-well plates to 80% confluence were infected with the different bacterial strains at a multiplicity of infection of 10. Infections were carried out in Hanks' balanced salt solution (HBSS) for 2 h at 37 °C. Non-adherent bacteria were removed from the monolayer by washing three times with HBSS. DMEM medium containing 100 µg ml⁻¹ gentamicin was then added to each well to kill extracellular bacteria. After 2 h incubation at 37 °C, monolayers were lysed with PBS containing 0.1% sodium deoxycholate and internalized bacteria were titred in L-agar plates. When required, EGF was added at the indicated concentration during the last 30 min of infection.

^{*} Percentage of bacteria added to the monolayer that survived the gentamicin treatment. Values are mean ± s.d. of triplicate samples.

† *S. typhimurium* wild type¹⁶.

‡ *S. typhimurium* *invA* mutant^{6,8}.

§ *E. coli* strain RDEC-1. This pathogenic strain of *E. coli* attaches readily (~50% of inoculum) to cultured Henle-407 cells (ref. 12, and our unpublished results).

Complementarity between sperm surface β -1,4-galactosyl-transferase and egg-coat ZP3 mediates sperm-egg binding

David J. Miller, Mary B. Macek & Barry D. Shur^{*}

Department of Biochemistry and Molecular Biology, The University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Boulevard, Box 117, Houston, Texas 77030, USA

DESPITE its importance, the molecular basis of mammalian gamete recognition has remained unclear. The enzyme β -1,4-galactosyltransferase (Gal-transferase) has been viewed traditionally as a biosynthetic component of the Golgi complex, but is also found on the surface of many cells where it can bind its specific glycoside substrate on adjacent cell surfaces or in the extracellular matrix^{1–3}. In mouse it has been suggested that Gal-transferase on the sperm head mediates fertilization by binding oligosaccharide residues in the egg coat, or zona pellucida^{4–9}, and that the ability of the zona pellucida to bind sperm is conferred by oligosaccharides of the ZP3 glycoprotein^{10,13}. However, it has

^{*} To whom correspondence should be addressed.

not been confirmed that Gal-transferase and ZP3 are in fact complementary gamete receptors whose interaction mediates sperm-egg binding. Here we show that mouse sperm Gal-transferase specifically recognizes those oligosaccharides on ZP3 that have sperm-binding activity, but does not interact with other zona pellucida glycoproteins. In contrast, all zona pellucida glycoproteins are recognized by non-sperm Gal-transferase, demonstrating a more stringent substrate specificity for the sperm enzyme. This interaction is required for sperm-egg binding because blocking or removing the binding site for Gal-transferase on ZP3 inhibits its ability to bind sperm. After the release of the sperm acrosome, the transferase relocates to a new membrane domain where it can no longer bind to ZP3, which is consistent with the inability of acrosome-reacted sperm to bind ZP3 or to initiate binding to the zona pellucida. Following fertilization, ZP3 is modified by egg cortical granule secretions so that it loses sperm receptor activity, which can be accounted for by a selective loss of its binding site for sperm Gal-transferase. These results show that sperm surface β -1,4-galactosyltransferase and the egg-coat glycoprotein ZP3 are complementary adhesion molecules that mediate primary gamete binding in the mouse.

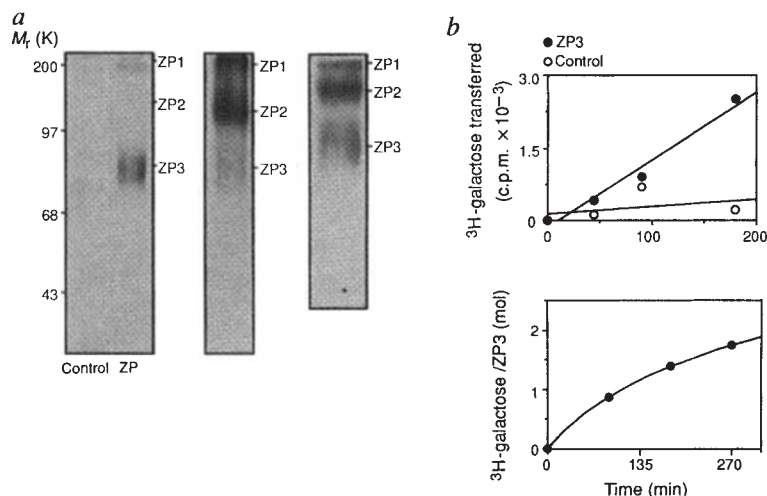
Although sperm surface Gal-transferase acts in a lectin-like capacity during sperm-egg binding by forming stable complexes with its *N*-acetylglucosamine-binding site in the zona pellucida (ZP)⁵⁻⁹, we took advantage of its catalytic properties to identify which of the three ZP glycoproteins interacts with Gal-transferase. ZP glycoproteins that bind the transferase were galactosylated with UDP-[³H]galactose and subsequently identified by SDS-polyacrylamide gel electrophoresis and fluorography. Only ZP3 was labelled with [³H]galactose (Fig. 1a). Surprisingly, when a biologically inappropriate Gal-transferase was used (affinity-purified bovine milk Gal-transferase), all three ZP glycoproteins were labelled (Fig. 1a), indicating that these all contained terminal *N*-acetylglucosamine, the monosaccharide recognized by Gal-transferase. Thus the membrane-bound

sperm Gal-transferase had a more selective substrate specificity, recognizing only ZP3; the presence of terminal *N*-acetylglucosamine was not in itself sufficient for binding sperm Gal-transferase. The possibility that the sperm enzyme might recognize other ZP glycoproteins subsequently lost as a result of proteolysis was excluded because ¹²⁵I-labelled ZP glycoproteins incubated under identical conditions still migrated in their correct positions. Purified ZP3 also served as a substrate for sperm Gal-transferase in quantitative assays (Fig. 1b) in which the [³H]galactose on ZP3 could be removed by β -1,4-specific galactosidase but not by α -galactosidase, confirming that the [³H]galactose was transferred by β -1,4-galactosyl-transferase (data not shown).

If Gal-transferase is indeed the sperm receptor for ZP3, then ZP3 should have several binding sites for Gal-transferase as ZP3 appears to induce the acrosome reaction by crosslinking its sperm receptor^{10,11}. When Gal-transferase binding sites were labelled by galactosylation, nonlinear regression estimated that ~2.6 galactose residues could be acquired by each molecule of ZP3 at saturation, thus allowing multivalent interactions (Fig. 1b). Consistent with this, aggregation of sperm Gal-transferase with multivalent antibodies induces the acrosome reaction, whereas monovalent reagents and Fab fragments do not¹².

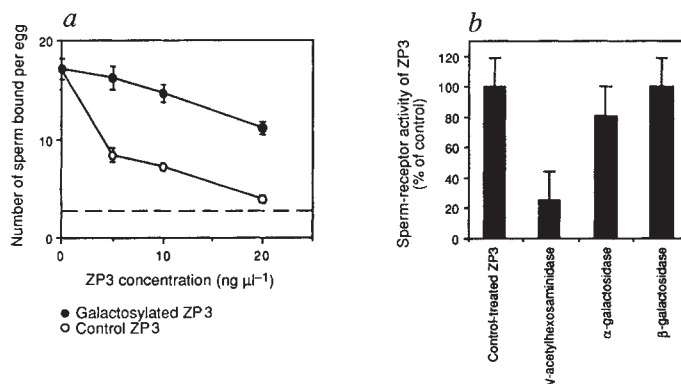
Partial characterization of the Gal-transferase-binding oligosaccharides on ZP3 show that they are serine/threonine-linked, like those with sperm receptor activity¹³. Removal of asparagine-linked oligosaccharides from ZP3 by *N*-glycanase digestion failed to release the Gal-transferase-binding oligosaccharides, whereas release of serine/threonine-linked oligosaccharides by mild alkaline hydrolysis (5 mM NaOH for 16 h at 37 °C) released 65% of the sperm-binding oligosaccharides (calculated from Fig. 2 in ref. 13) and concurrently released 50% of the Gal-transferase-labelled oligosaccharides, which were of similar size to those with sperm-binding activity (~4K). Under these conditions *N*-glycanase efficiently removed

FIG. 1 a, Sperm Gal-transferase selectively recognizes oligosaccharides on ZP3 from unfertilized eggs, although non-sperm Gal-transferase binds all ZP glycoproteins, thus demonstrating a more stringent substrate specificity for the sperm Gal-transferase. Galactosylation of ZP glycoproteins requires that either Gal-transferase or its glycoprotein substrate is soluble in order to produce maximal signal. Therefore, in the left panel, intact sperm were used as the biologically relevant source of Gal-transferase with heat-solubilized ZP, which retains sperm-receptor activity²⁰. CF1 cauda epididymal sperm were capacitated in KRBT³⁶, washed and 2×10^6 sperm were incubated at 37 °C in 100 μ l containing 100 μ M UDP-[³H]Gal (DuPont-NEN) ($10.5 \text{ Ci mmol}^{-1}$), 10 mM MnCl₂, 1 mM 5'AMP (competitive inhibitor of UDP-Gal degradation)¹⁴, 100 ng ml⁻¹ pertussis toxin (acrosome reaction inhibitor)³⁷, and 80 μ g heat-solubilized whole ZP or no ZP (control). ZP were purified by density-gradient centrifugation of ovarian homogenates¹⁵. After 3 h, sperm were removed by centrifugation and the supernatant analysed by SDS-PAGE and fluorography. *M_r* standards are shown to the left (in thousands). Results are representative of three experiments. To verify that the absence of labelling of ZP1 and ZP2 was not due to proteolysis, ¹²⁵I-labelled ZP were incubated with sperm and unlabelled UDP-Gal under identical conditions: results are shown in the middle panel. The right panel shows that soluble non-sperm Gal-transferase has a less selective substrate specificity than does membrane-bound sperm Gal-transferase. Soluble Gal-transferase (25 μ g from bovine milk; Sigma) was affinity-purified⁹ and incubated with 160 structurally intact ZP (mechanically isolated from oviductal eggs) in 400 μ l HEPES-buffered saline with 30 μ M UDP-[³H]Gal and 10 mM MnCl₂ for 6 h. The ZP glycoproteins were analysed by SDS-PAGE and fluorography. Results are representative of four experiments. b, ZP3 has multiple binding sites for sperm Gal-transferase. The initial kinetics of sperm Gal-transferase glycosylation of ZP3 are shown in the upper panel. Capacitated sperm (2×10^6) were incubated with 20 μ g ZP3 (●) or BSA (controls; ○) and 30 μ M UDP-[³H]Gal as in a, and aliquots of the supernatant containing labelled ZP3 were removed at the times



indicated and subjected to high-voltage borate electrophoresis⁹. In the lower panel are shown the results used to estimate the total number of binding sites on ZP3 for Gal-transferase. Sperm (3×10^6) were incubated with 10 μ g ZP3 in 200 μ l KRBT³⁶ in the presence of 1 mM PUGNAC (*N*-acetylglucosamino-1,5-lactone oxime) (CarboGen, Zurich), an *N*-acetylglucosaminidase inhibitor³⁸. Results (c.p.m.) from controls without ZP3 were subtracted from samples containing ZP3. Logistic fit nonlinear regression was used to estimate the maximum amount of [³H]galactose transferred to ZP3. To determine whether the galactose transferred was due to β 1,4-Gal-transferase or other galactosyltransferases, products were digested with 100 mU per ml of β -1,4-specific galactosidase (*Diplococcus*; Boehringer-Mannheim) or α -galactosidase (*E. coli*; Boehringer-Mannheim) for 16 h at 37 °C, pH 6, and analysed by high-voltage borate electrophoresis (data not shown). For controls, inactive glycosidases were used.

FIG. 2 Either blocking (a) or removing (b) the Gal-transferase-binding site on soluble ZP3 reduces its ability to inhibit sperm-egg binding competitively. **a**, ZP3 (10 μg) was incubated with 0.2 mg ml⁻¹ affinity-purified bovine milk Gal-transferase and 1.25 mM UDP-Gal in HEPES-buffered saline with 10 mM MnCl₂ for 12 h at 37 °C to block (by galactosylation) the terminal *N*-acetylglucosamine recognized by Gal-transferase (●). Controls (○) used heat-inactivated enzyme. The products were heat-treated to inactivate Gal-transferase and dialysed to remove UDP-Gal, both of which inhibit sperm-egg binding^{5,8,9}. Recovery was monitored by addition of ¹²⁵I-labelled ZP. To determine whether galactosylated ZP3 or mock-incubated ZP3 could inhibit sperm-egg binding, each was added to sperm and eggs in competitive binding assays^{8,9,19}. Results are averages of 3 experiments using duplicate droplets. Two-cell embryos were included in each droplet to determine nonspecific binding and averaged ~3 sperm per embryo, which is indicated by the dotted line. Data were normalized by logarithmic transformation and evaluated by analysis of variance³⁹. Means were compared using the Studentized range test and standard errors are indicated by bars. **b**, Sperm receptor activity of glycosidase-treated ZP3 measured in competitive binding assays. Soluble ZP3 (5 μg) was incubated with 100 mU per ml *N*-acetylhexosaminidase (Jack Bean, from Sigma; specific activity towards *N*-acetylglucosamine twice that of *N*-acetylgalactosamine), α -galactosidase (*E. coli*; Boehringer-Mannheim) or β -galactosidase (*Diplococcus*; Boehringer-Mannheim) at 37 °C for 6 h in citrate-phosphate buffer with a protease inhibitor cocktail at pH 5, 6 or 6, respectively. No contaminating exoglycosidase activities were detected using *p*-nitrophenyl substrates (limit



of detection was 5 μU of activity). Products were heat-inactivated and 12.5 ng μl^{-1} treated ZP3 was added to sperm and eggs in competition assays. Soluble ZP3 treated with heat-inactivated glycosidases (controls) reduced the number of sperm bound to eggs by 14 sperm (27.8 sperm per egg without ZP3; 13.8 sperm per egg with inactive-enzyme-treated ZP3), a value defined as 100% sperm-receptor activity of soluble ZP3. Results are averages of 3 experiments using duplicate droplets. Data were compared by analysis of variance and individual means by the Studentized range test. Bars indicate standard errors.

asparagine-linked chains, and there was no evidence of proteolysis during alkali digestion as determined by SDS-polyacrylamide gel electrophoresis (data not shown).

Although sperm Gal-transferase bound selectively to ZP3 oligosaccharides, it was not clear whether this interaction was required for sperm binding to the zona pellucida. To assess this directly, we blocked Gal-transferase-binding sites in soluble ZP3 by incubation of purified ZP3 with UDP-galactose and soluble

affinity-purified Gal-transferase, and the resulting sperm receptor activity of this galactosylated ZP3 was assayed by competition. Control ZP3 incubated with heat-inactivated transferase competitively reduced sperm-egg binding to almost background levels, but galactosylated ZP3, which was no longer able to bind Gal-transferase, had significantly less sperm-receptor activity and was less effective as a competitive inhibitor of sperm-egg binding (Fig. 2a).

FIG. 3 After the acrosome reaction, Gal-transferase is redistributed to the lateral sperm head where it loses affinity for ZP3 (a), and has no affinity for other ZP glycoproteins (b). Following fertilization, ZP3 is no longer recognized by sperm Gal-transferase (c), although it is still recognized by non-sperm Gal-transferase (d). **a**, Capacitated sperm were treated with 10 μM divalent cation ionophore A23187 for 60 min, washed, and incubated with ZP3 and UDP-[³H]Gal as for Fig. 1. Frequencies of acrosome reactions were determined by HS21-antibody staining⁴. In the average of three experiments, 81% of the control sperm (no ionophore) had intact acrosomes at t_0 , which decreased to 66% by t_{180} (●). Ionophore-treated sperm had 32% intact acrosomes at t_0 , which decreased to 27% by t_{180} (○). The same experiment was performed in the presence of 1 mM PUGNAC (see legend to Fig. 1) to inhibit any released acrosomal *N*-acetylglucosaminidase³⁸. Although total c.p.m. increased owing to inhibition of *N*-acetylglucosaminidase activity, the relative difference between acrosome-intact (●) and acrosome-reacted (○) sperm remained unchanged. **b**, Control sperm or ionophore-treated sperm were incubated with 80 μg of ZP and UDP-[³H]Gal as for Fig. 1. Left panel, control sperm had 67% intact acrosomes at t_0 , which decreased to 61% by t_{180} . Ionophore-treated sperm had 23% intact acrosomes at t_0 , which decreased to 14% by t_{180} . ZP glycoproteins were analysed by SDS-PAGE and fluorography. The gel is representative of two experiments. Right panel, to test for ZP proteolysis under these conditions, sperm were incubated with [¹²⁵I]-ZP as in Fig. 1a. **c**, 2,000 ZP from unfertilized oviductal eggs (●) or two-cell embryos (○) were mechanically isolated, heat-solubilized, and incubated with sperm under conditions identical to those described for Fig. 1. Background c.p.m. from controls without ZP were subtracted from samples containing ZP. **d**, Structurally intact ZP from 160 oviductal unfertilized eggs (Unfert) or two-cell embryos (Fert) were incubated with 25 μg affinity-purified bovine milk β -1,4-Gal-transferase and UDP-[³H]Gal, as for Fig. 1, for 6 h. The [³H]galactose-labelled ZP glycoproteins were analysed by SDS-PAGE and fluorography.

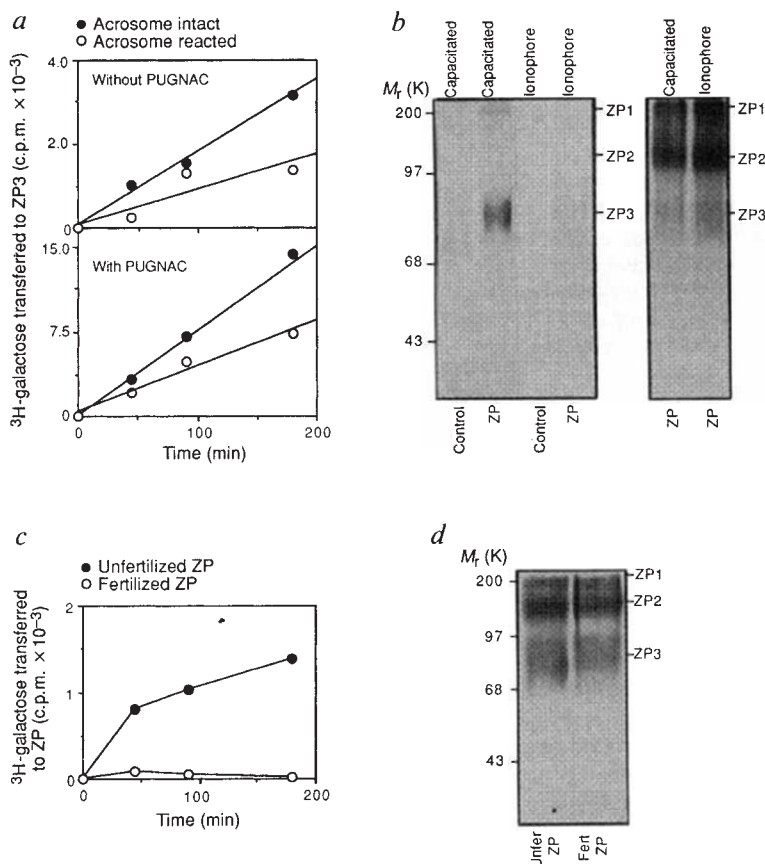
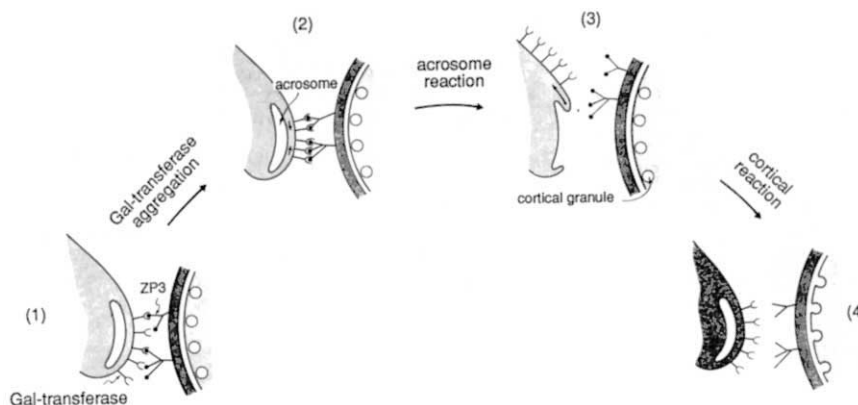


FIG. 4 Model for the function of sperm Gal-transferase and egg ZP3 during sperm-egg binding. Initial sperm-egg recognition (1) is mediated by sperm Gal-transferase binding to terminal *N*-acetylglucosamine residues (●) on ZP3 oligosaccharides; the multivalent ZP3 binds 2–3 Gal-transferase moieties, thereby aggregating the transferase (2); aggregation of Gal-transferase initiates the acrosome reaction (3), during which time the plasma membrane overlying the acrosome is released, while Gal-transferase relocates to the lateral aspect of the sperm head. After the acrosome reaction, the fertilizing sperm binds to the egg in an interaction that does not apparently involve ZP3 and Gal-transferase, and which probably requires unique adhesion components. (4) The fertilizing sperm penetrates the ZP and fuses with the underlying egg plasma membrane. Fusion activates the egg, resulting in the exocytosis of egg cortical granules that contain several enzymes, possibly including *N*-acetylglucosaminidase.



N-acetylglucosaminidase acts on ZP3 to remove terminal *N*-acetylglucosamine residues from the sperm-receptor oligosaccharides, which prevents binding of accessory sperm to the ZP.

If Gal-transferase-binding sites of ZP3 were removed (92% removed; $P < 0.05$) by digestion with *N*-acetylhexosaminidase, ZP3 lost its sperm-receptor activity (75% lost; $P < 0.05$, Fig. 2b). In parallel assays, α -galactosidase had no effect on ZP3 sperm-receptor activity ($P > 0.1$), in contrast to other reports¹⁴, nor did β -galactosidase (Fig. 2b). α -Galactosidase was active and efficiently removed terminal α -galactose residues from ZP3, because α -1,3-galactosyltransferase sites were created by pre-treatment with α -galactosidase (160% of controls; $P < 0.05$).

Following the acrosome reaction, sperm show lower affinity for ZP3 but an increased affinity for ZP2, and a concomitant relocalization of Gal-transferase to a new membrane domain^{4,15,16}. Consistent with this, Gal-transferase on acrosome-reacted sperm no longer recognized ZP3 (a 50% reduction in acrosome-intact sperm following ionophore treatment gave a 50% reduction in Gal-transferase labelling of ZP3 (Fig. 3a); less than 20% acrosome-intact sperm resulted in barely detectable ZP3 labelling (Fig. 3b)). The reduced labelling was not due to released acrosomal *N*-acetylglucosaminidase, because similar results were seen in the presence of the *N*-acetylglucosaminidase inhibitor, *N*-acetylglucosamino-1,5-lactone oxime (PUGNAC; Fig. 3a). The residual binding of Gal-transferase to ZP3 seen with ionophore-treated sperm is probably due to the acrosome-intact sperm remaining in the population, but may also reflect a decreased affinity for ZP3 by the transferase on acrosome-reacted sperm. Nevertheless, relocalization of the Gal-transferase to a new membrane domain after the acrosome reaction results in reduced ZP3 binding. Surprisingly, Gal-transferase on acrosome-reacted sperm still retains normal activity towards other substrates, indicating that the reduced binding to ZP3 is specific (acrosome-intact/acrosome-reacted sperm: 4,637/4,921 c.p.m. [³H]galactose transferred to *N*-acetylglucosamine per 10⁶ sperm per hour; 166/251 c.p.m. [³H]galactose transferred to ovalbumin per 10⁶ sperm per hour). Finally, there was no binding of Gal-transferase on acrosome-reacted sperm to ZP2 or ZP1 (Fig. 3b), suggesting that the enzyme is not involved in secondary binding of acrosome-reacted sperm to either of these glycoproteins, events that require the activation of unique sperm-surface receptors^{16–18}. No proteolysis of ¹²⁵I-labelled ZP glycoproteins incubated with sperm was detected.

After egg activation, cortical granule secretions destroy ZP3 sperm-binding activity, thus preventing polyspermy^{19,20}. Zona pellucida from fertilized eggs could no longer serve as substrate for Gal-transferase (Fig. 3c), showing that the loss of ZP3 sperm-receptor activity is correlated with a loss of recognition by sperm Gal-transferase. As this transferase requires terminal

N-acetylglucosamine residues on ZP3 for binding, and cortical granules in eggs of some species, including mouse (unpublished observations), contain *N*-acetylglucosaminidase as the principal glycosidase^{21,22}, the release of this enzyme at fertilization might inactivate ZP3 by removing its Gal-transferase-binding site. Zona pellucida from fertilized eggs still have abundant substrate for non-sperm Gal-transferase (Fig. 3d), suggesting that any putative cortical granule *N*-acetylglucosaminidase must be specific for oligosaccharides recognized by sperm Gal-transferase. Alternatively, other sperm Gal-transferase recognition determinants may be removed by cortical granule enzymes.

Results presented here and elsewhere^{5,7–9} show that sperm Gal-transferase fulfils several criteria for a ZP3 receptor (Fig. 4); it behaves as an integral plasma membrane component overlying the intact acrosome; blocking its activity inhibits sperm-egg binding; it selectively recognizes oligosaccharides on ZP3; the sperm-receptor activity of ZP3 depends on its Gal-transferase binding site; on acrosome-reacted sperm Gal-transferase no longer binds ZP3; and its binding site on ZP3 is lost after fertilization. Although there are other sperm proteins reported to bind ZP3, it is not yet clear whether they are plasma membrane proteins that satisfy the criteria for a ZP3 receptor, nor if they function during sperm-egg binding^{23,24}.

The function of β -1,4-galactosyltransferase during fertilization may not be limited to the mouse as it has been found on all mammalian sperm tested so far^{25–27} and is confined to the plasma membrane in those species of sperm analysed by subcellular fractionation⁶. In some other species, blocking sperm Gal-transferase or destroying its binding site on the zona pellucida also reduces sperm-egg binding^{28–30}. Although the basis of species-specific fertilization is unknown, studies in sea urchins indicate that both carbohydrate and protein determinants may be involved³¹. Similarly, species specificity in mammals could be mediated by Gal-transferase recognition of both carbohydrate and protein determinants of ZP3, as the substrate specificity of many glycosyltransferases depends on oligosaccharide and protein components, and not just on a simple terminal monosaccharide^{32–35}. In any case, mammalian gamete recognition could well be dictated by a multiplicity of adhesion molecules that collectively impart species specificity^{16–18,23,24}. □

Received 19 March; accepted 8 May 1992.

- Shur, B. D. *Glycobiology* **1**, 563–575 (1991).
- Roth, S. Q. *Rev. Biol.* **48**, 541–563 (1973).
- Roseman, S. *Chem. Phys. Lipids* **5**, 270–297 (1970).
- Lopez, L. C. & Shur, B. D. *J. Cell Biol.* **105**, 1663–1670 (1987).
- Shur, B. D. & Neely, C. A. *J. Biol. Chem.* **263**, 17706–17714 (1988).

6. Fayer-Hosken, R. A., Caudle, A. B. & Shur, B. D. *Molec. Reprod. Dev.* **28**, 74-78 (1991).
7. Benau, D. A. & Storey, B. T. *Biol. Reprod.* **39**, 235-244 (1988).
8. Shur, B. D. & Hall, N. G. *J. Cell Biol.* **95**, 574-579 (1982).
9. Lopez, L. C. *et al.* *J. Cell Biol.* **101**, 1501-1510 (1985).
10. Florman, H. M., Bechtol, K. B. & Wassarman, P. M. *Dev. Biol.* **106**, 243-255 (1984).
11. Leyton, L. & Saling, P. *J. Cell Biol.* **108**, 2163-2168 (1989).
12. Macek, M. B., Lopez, L. C. & Shur, B. D. *Dev. Biol.* **147**, 440-444 (1991).
13. Florman, H. M. & Wassarman, P. M. *Cell* **41**, 313-324 (1985).
14. Bleil, J. D. & Wassarman, P. M. *Proc. natn. Acad. Sci. U.S.A.* **85**, 6778-6782 (1988).
15. Bleil, J. D. & Wassarman, P. M. *J. Cell Biol.* **102**, 1363-1371 (1986).
16. Bleil, J. D., Greve, J. M. & Wassarman, P. M. *Dev. Biol.* **128**, 376-385 (1988).
17. Blobel, C. P. *et al.* *Nature* **356**, 248-252 (1992).
18. Jones, R., Brown, C. R. & Lancaster, R. T. *Development* **102**, 781-792 (1988).
19. Bleil, J. D. & Wassarman, P. M. *Cell* **20**, 873-882 (1980).
20. Wassarman, P. A. *Rev. Biochem.* **57**, 415-442 (1988).
21. Lambert, C. C. *Development* **105**, 415-420 (1989).
22. Prody, G. A., Greva, L. C. & Hedrick, J. L. *J. exp. Zool.* **235**, 335-340 (1985).
23. Bleil, J. D. & Wassarman, P. M. *Proc. natn. Acad. Sci. U.S.A.* **87**, 5563-5567 (1990).
24. Leyton, L. & Saling, P. *Cell* **57**, 1123-1130 (1989).
25. Humphreys-Beher, M. G., Garrison, P. W. & Blackwell, R. E. *Fertil. Steril.* **54**, 133-137 (1990).
26. Miller, D. J., Cross, N. L., Vazquez-Levin, M. & Shur, B. D. in *Comparative Spermatology-20 Years After* (eds Baccetti, B.) 569-574 (Serono Symp., Raven, New York, 1992).
27. Sullivan, R., Ross, P. & Berube, B. *Biochem. biophys. Res. Commun.* **162**, 184-188 (1989).
28. Berger, T., Davis, A., Wardrip, N. J. & Hedrick, J. L. *J. Reprod. Fertil.* **86**, 559-565 (1989).
29. Fayer-Hosken, R. A., McBride, C. E. & Brackett, B. G. *Biol. Reprod.* **36**, 53 (1987).
30. Mori, E. *et al.* *Biochemistry* **30**, 2078-2087 (1991).
31. Ruiz-Bravo, N. & Lennarz, W. J. *Dev. Biol.* **118**, 202-208 (1986).
32. Baranski, T. J., Faust, P. L. & Kornfeld, S. *Cell* **63**, 281-291 (1990).
33. Begovac, P. C., Hall, D. E. & Shur, B. D. *J. Cell Biol.* **113**, 637-644 (1991).
34. Passaniti, A. & Hart, G. W. *Cancer Res.* **50**, 7261-7271 (1990).
35. Schachter, H. *Biochem. Cell Biol.* **64**, 163-181 (1986).
36. Neill, J. M. & Olds-Clarke, P. *Gamete Res.* **20**, 459-473 (1988).
37. Endo, Y., Lee, M. A. & Kopf, G. S. *Dev. Biol.* **119**, 210-216 (1987).
38. Beer, D. & Vasella, A. *Helv. chim. Acta* **68**, 2254-2274 (1985).
39. Neter, J. & Wasserman, W. *Applied Linear Statistical Models* (Irwin, Homewood, Illinois, 1974).

ACKNOWLEDGEMENTS. We thank D. Wolf for HS-21 monoclonal antibody, X. Gong for technical assistance, F. Mao for help with nonlinear curve fitting, A. Tizenor for computer graphics, and members of our laboratory for comments. This research was supported by grants from the NIH (B.D.S.), a National Research Service Award (D.J.M.) and a Cancer Centre (Core) support grant. This letter is dedicated to the memory of Mary Beth Macek.

detected inside cells permeabilized with detergent. The lace-like staining pattern suggested that CD3 ϵ was located in the ER⁵. Transfection of mutant 11, which lacked amino acids 171 to 180, resulted in the expression at the plasma membrane of the polypeptide as revealed by staining of intact cells (Fig. 2). The intracellular staining pattern showed that the polypeptide was located predominantly in the Golgi apparatus rather than in the ER. These results indicate that the mutant-11 polypeptide probably escaped retention in the ER and was exported to the plasma membrane. Deletion of the transmembrane region (deletion 13) also prevented ER retention. This membrane-anchor-free CD3 ϵ polypeptide was located in the Golgi apparatus but was not detected on the surface of intact cells (Fig. 2). It was, however, secreted to the extracellular medium (data not shown). Deletion of the last five amino acids of the C terminus of CD3 ϵ (deletion 12) resulted in partial loss of ER retention, the plasma membrane of transfected cells being only weakly stained (data not shown).

To verify whether the sequence deleted in mutant 11 contained an ER retention motif, a chimaeric protein was constructed in which this KGQRDLVSL sequence (one-letter notation) was attached to the C-terminal end of human CD4 (CD4 ϵ 10; Table 1). Transfection of wild-type CD4 in COS cells resulted in expression of the polypeptide on the plasma membrane and inside the cell in the Golgi apparatus (Fig. 3a). In contrast, CD4 ϵ 10 was retained in the ER and not detected on the cell surface. The sensitivity of chimaeric CD4 ϵ 10 protein to endoglycosidase H (endo H) was measured to verify that this protein had not passed through the Golgi apparatus. As shown in Fig. 3b, 68% of wild-type CD4 was converted to a partially endo H-resistant form after one hour of metabolic labelling with

An endoplasmic reticulum retention signal in the CD3 ϵ chain of the T-cell receptor

Arrate Mallabiarrena, Manuel Fresno & Balbino Alarcón*

Centro de Biología Molecular, Universidad Autónoma de Madrid, CSIC, Cantoblanco, Madrid 28049, Spain

ISOLATED polypeptide chains of the T-cell antigen receptor complex are degraded or retained in the endoplasmic reticulum (ER)¹⁻⁴. Assembly of the multisubunit complex allows the individual chains to escape retention in the ER and to be expressed on the cell surface. We engineered a series of deletions in the CD3 ϵ subunit of the human T-cell receptor in order to find the sequences responsible for its retention in the ER. Deletion of amino acids 171 to 180 in the cytosolic tail resulted in the cell-surface expression of the isolated chain. This sequence also promotes retention when it is appended to CD4, a plasma membrane protein. Mutagenesis of the 10-amino-acid CD3 ϵ sequence established that the tyrosine and serine residues are important for ER retention. This and other ER retention signals must be hidden when a complete T-cell receptor complex is assembled in order to allow its expression on the cell surface.

Assuming that the putative ER retention signal of CD3 ϵ is conserved among different species, we generated a series of 13 deletion mutants to encompass the most homologous regions. These deletions affected 124 of the 185 amino acids of the human CD3 ϵ sequence (Fig. 1). The effects of each deletion were tested on the cell-surface expression and intracellular distribution of the CD3 ϵ polypeptide in COS cells transfected with the mutant constructs. As shown in Fig. 2, wild-type CD3 ϵ was not expressed on the surface of intact COS cells although it was

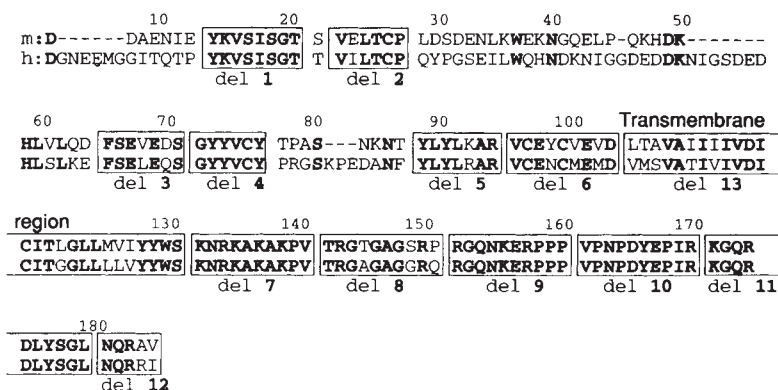
TABLE 1 Mutational analysis of the ER retention sequence

Sequence	Construct	Intracellular pattern	Surface expression
EPKKGQRDLVSLNQRR	CD3 ϵ	ER	—
EPKRNQRR	CD3 ϵ del 11	Golgi	+++
EPKKGQRDLVSL	CD3 ϵ del 12	Golgi/ER	++
CSPI	CD4	Golgi	+++
GSPIKGQRDLVSLNQRR	CD4 ϵ 15	ER	—
CSPIKGQRDLVSLWQT	CD4 ϵ 10	ER	—
CSPIKGQRDLVSL	CD4 ϵ 10t	Golgi	—
CSPINQRR	CD4 ϵ 5h	Golgi	+++
CSPIQRAV	CD4 ϵ 5m	Golgi	+++
CSPIKHGDLVSLWQT	CD4 ϵ 10	Golgi	+++
CSPIKGQRV	CD4mt32	Golgi	+++
CSPIQRDLWQT	CD4 ϵ 4	Golgi	+++
CSPIDVSLWQT	CD4 ϵ 6	ER/Golgi	—
CSPIKGQRDLVSLWQT	CD4mt12	ER/Golgi	++
CSPIKGQRDLVSLWQT	CD4mt6	ER	—
CSPIKGQRDLVSLWQT	CD4mt18	Golgi	+++
CSPIKGQRDLVSLWQT	CD4mt24	ER	—
CSPIKGQDLVSLWQT	CD4mt34	ER	—

The carboxy-terminal acid sequence (in single-letter code) of the constructs is represented as small capital letters for invariant portions and as capital letters for the retention sequence or its variants. Amino-acid changes in the 10-amino-acid retention sequence are shown in bold type. Minus signs represent absence of expression; +++, strong reaction; ++, a weak reaction. Surface expression, retention in the ER and expression in the Golgi apparatus were evaluated as described for Fig. 2. CD4 ϵ 10, CD4 ϵ 4 and CD4 ϵ 6 were obtained by polymerase chain reaction (PCR) using oligonucleotide 1 as a 5' primer (Fig. 3 legend) and as 3' primers the oligonucleotides (G)₆ATCCACAGCCCTGGTACAGCCCGCTGGCCCTTAATGGGGCTACATGTCT, (G)₆ATCCACAGGTCCTGGCTGAATGGGGCTACATGTCT, and (G)₆ATCCACAGCCAGAATACAGGTCAGTGGGGCTACATGTCT. CD4 ϵ 10t, CD4 ϵ 5h and CD4 ϵ 5m were constructed by PCR using oligonucleotide 1 as 5' primer and oligonucleotides (G)₆ATCCACAGGCC AGAATACAGGTCCTGGCTGGCCCTTAATGGGGCTACATGTCT, (G)₆ATCCACAGTGGCTCTCTGATTAATGGGGCTACATGTCT, and (G)₆ATCCCTCAGTGGCTCTCTGATTAATGGGGCTACATGTCT as 3' primers. CD4mt32, mt12, mt6, mt18, mt24 and mt34 were obtained by PCR using oligonucleotide 1 as the 5' primer and as a 3' primer the degenerate oligonucleotide (G)₆ATCCAXAAGCXATAXAAGTXXCGTGGXXCTTAATGGGGCTACATGTCT, where X is a mixture of 90% C and 3.3% each of A, T and G. Z is a mixture of 51.6%T, 45% of C and 1.6% each of G and A. This oligonucleotide mutates each position of the decapeptide KGQRDLVSL with a frequency of ~10%, except for Y, which is mutated in 47% of cases. After ligation of the PCR product into the pSR α vector, insert-positive colonies were sequenced and transfected into COS cells for analysis.

* To whom correspondence should be addressed.

FIG. 1 Map of deletions of human CD3 ϵ showing the amino acid sequence (in one-letter code) of mature human (h) and murine (m) CD3 ϵ chains. Conserved amino acids are shown in bold type. The positions of human CD3 ϵ amino acids are indicated above the murine sequence. Regions deleted in human CD3 ϵ together with the homologous murine regions are boxed. Deletion 13 involves all the transmembrane region of human CD3 ϵ . Sequences were aligned according to ref. 2. METHODS. Deletional loop mutagenesis was achieved in human CD3 ϵ as described¹⁷. In brief, 40 μ g vector pSR α (ref. 18) and 40 μ g pSR α -CD3- ϵ (ref. 19) were cut, respectively, with *Xho*I and *Xmn*I, and subsequently treated with alkaline phosphatase and precipitated with ethanol. Both DNA pellets were dried, redissolved in 25 μ l water, mixed, denatured by addition of 50 μ l 0.4 M NaOH and incubated at room temperature for 10 min. Tris-HCl (900 μ l, 0.1 M) at pH 7.5 was added dropwise to neutralize the pH, and reannealing was allowed at 68 $^{\circ}$ C for 2 h. The sample was slowly cooled to room temperature and precipitated with ethanol. Pellets were redissolved in 250 μ l 0.1 M Tris-HCl, pH 7.6, then 45 μ l was mixed with 5 μ l (50 pmol) of mutagenic 5'-phosphorylated oligonucleotide and 1 μ l each of 100 mM MgCl₂, 10 mM ATP, a mixture containing each deoxyribonucleotide at 5 mM, and 200 mM DTT. Two units of large fragment (Klenow) DNA polymerase I and 100 U of T4 DNA ligase were added and the mixture incubated overnight at room temperature. DNA was precipitated with ethanol, redissolved in 10 μ l 10 mM



Tris-HCl, pH 8.0, 1 mM EDTA, of which 5 μ l was used to transform competent *E. coli* DH5 α . Bacteria were seeded on nitrocellulose filters. Replica filters were hybridized with 5' ³²P-labelled mutagenic oligonucleotide under stringent conditions (60 $^{\circ}$ C in 6 $\times$ SSC, 0.1% SDS) to detect colonies containing the deleted sequence. Deletions were confirmed by DNA sequencing. The mutagenic oligonucleotides used were 24-nucleotide oligomers whose sequences hybridized with the first 12 nucleotides upstream and with the first 12 nucleotides downstream from the sequence to be deleted.

³⁵S-methionine plus ³⁵S-cysteine, and was completely converted after eight hours. None of the CD4 ϵ 10 chimaera was processed to a mature form, however, even after an 8-hour chase. The lack of maturation of CD4 ϵ 10 in the Golgi did not result from an increased susceptibility to ER degradation because CD4 ϵ 10 was degraded at a slower rate than wild-type CD4 (52% versus 69% after 8 h). These results confirmed that the CD4 ϵ 10 chimaera had not been processed in the Golgi apparatus and therefore that region 11 of CD3 ϵ contains an ER retention signal. The CD4 ϵ 15 chimaera, resulting from the addition of the last 15 amino acids of CD3 ϵ (regions 11 plus 12) to CD4, was also retained in the ER (Table 1). But addition of the last five amino acids of human or murine CD3 ϵ , missing in deletion 12, to the C terminus of CD4 (constructs CD4 ϵ 5 and CD4 ϵ 5m) did not

promote ER retention of the protein. These results suggest that the C-terminal sequences NQRRRI and NQRAV of human and murine CD3 ϵ do not contain a retention signal. The low surface expression of mutant 12 also indicated that the distance of the motif in region 11 from the C terminus was important for its ER retention function. Consistent with this idea, CD4 ϵ 10t, which had a stop codon immediately following the KGQRDLYSGL sequence, was not retained (Table 1).

Sequence analysis revealed that six positions of the ER retention motif of CD3 ϵ were conserved in the unrelated polypeptide, CD3 ζ (ref. 6). This region in human CD3 ζ corresponded to amino acids 135–145 (KGHDGLYQGL), and comprises part of an 18-residue motif involved in signal transduction⁷. A chimaeric protein containing the ten amino acids from ζ

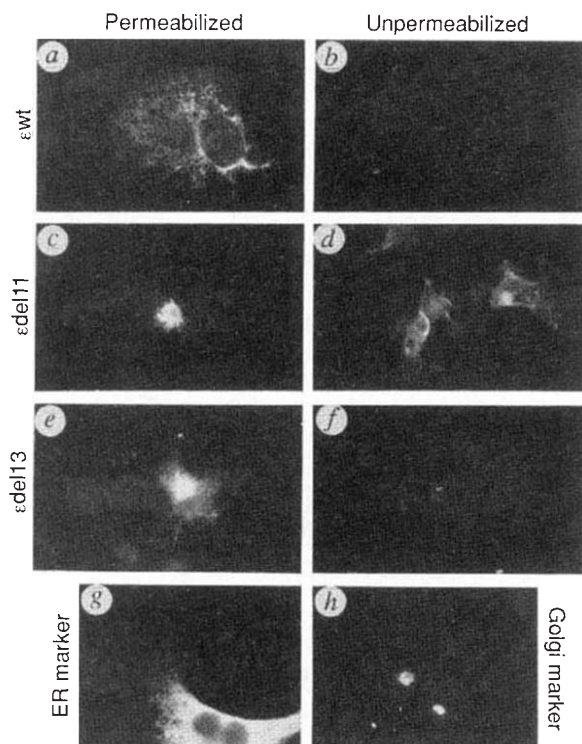
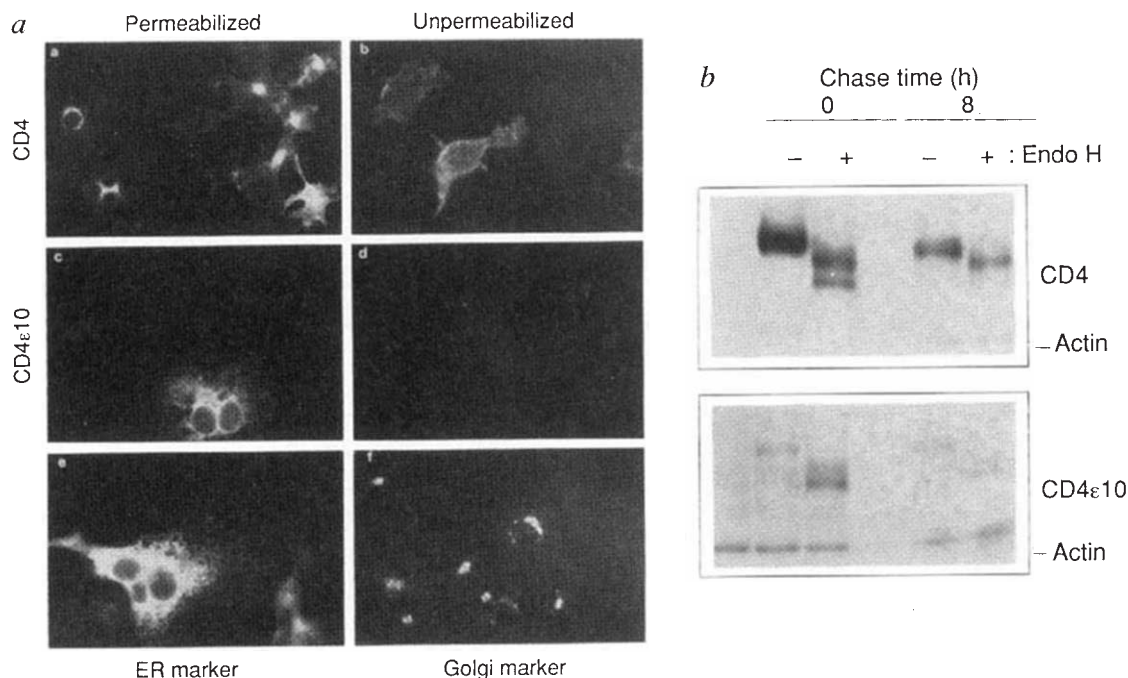


FIG. 2 Immunofluorescence staining of COS cells transfected with CD3 ϵ constructs. COS-7 cells were transfected with an expression vector encoding wild-type CD3 ϵ cDNA (ϵ wt), CD3 ϵ deleted of region 11 (ϵ del11) or CD3 ϵ deleted of region 13 (ϵ del13). Cells were fixed with 2% paraformaldehyde 72 h after transfection and either permeabilized with 0.1% saponin or left unpermeabilized. Permeabilized cells were stained with the CD3 ϵ antibody APA1/1, which probably recognizes an internal epitope of CD3 ϵ . Unpermeabilized cells were stained with SP34 antibody which recognizes an extracellular epitope of CD3 ϵ . All photographs were taken at a $\times$ 630 magnification in a Zeiss Axiophot microscope. Transfection with a complementary DNA construct that expresses a chimaera of CD8 and adenoviral E3/19K protein was used as an ER marker¹¹. C6-NBD-ceramide (*N*-(7-nitrobenz-2-oxa-1,3 diazol-4-yl aminocaproyl) sphingosine) was used as a Golgi marker²⁰. Transfection and staining were as described²¹.

FIG. 3 Retention of a tagged CD4 protein. **a**, Immunofluorescence staining of COS cells transfected with CD4 constructs. COS-7 cells were transfected with 10 μ g pSR α -CD4 which encodes wild-type human CD4, or with 10 μ g pSR α -CD4 ϵ 10 which encodes a complete human CD4 protein with peptide KGQRDLYSGLWIQT appended to the C terminus. Cells were fixed 72 h after transfection as for Fig. 2 and stained with the HP2/4 CD4 monoclonal antibody²². The WIQT sequence was added to the retention decapeptide to distance it from the C-terminal end. Magnification $\times 400$, in **a**, $\times 630$ in **b-f**. **b**, Endoglycosidase H (endo H) sensitivity of oligosaccharides attached to CD4 and CD4 ϵ 10. CD4 immunoprecipitates from pulse-chase labelled transfected COS cells were treated with endo H (+) or left untreated (-) before they were electrophoresed. Position of contaminant actin band is indicated.



METHODS. pSR α -CD4 was obtained by removing a 1.8 kb cDNA insert from the pSP65-CD4 vector with *Eco*RI and *Bam*HI and ligating unidirectionally into the same sites of Bluescript (Stratagene). The Bluescript CD4 construct was then cut with *Xho*I and *Bam*HI and ligated into a previously cut pSR α vector. The CD4 ϵ 10 chimera was obtained by PCR using the 5' primer (C)₆TCGAGCAAGGCCACAATGAACCGG (oligonucleotide 1), which corresponds to the N-terminal region of CD4 including the initiation codon and the 3' primer (G)₆ATCCACAGGCCAGAAATACAGGTCCTGGCCTTTAATGGGGCTAC-ATGTCT, which anneals to the last coding nucleotides of CD4 and introduces amino acids KGQRDLYSGLWIQT to the C-terminal tail of CD4. Amino acids WIQT were added to keep the retention sequence away from the C terminus

and were encoded in the 3' primer (W) and in the vector itself (IQT). The 5' and 3' primers contained *Xho*I and *Bam*HI sites, respectively, for unidirectional cloning into the pSR α vector. The PCR product was cut, ligated into pSR α and the ligation mix used to transform DH5 α -competent *E. coli*. A positive colony was sequenced to verify the sequence and was subsequently used for transfections. To measure maturation of the CD4 and CD4 ϵ 10 oligosaccharides, COS cells were transfected with 10 μ g of each construct and labelled for 1 h with 1 mCi of ³⁵S-methionine plus ³⁵S-cysteine. Plates were immediately lysed in immunoprecipitation buffer or incubated for 8 h at 37 °C with fresh medium. Immunoprecipitation with the CD4 antibody HP2/4 and endoglycosidase H treatment were as described²³. The lower amount of CD4 ϵ 10 protein compared with CD4wt in **b** was not normally seen and could be an artefact of preparation.

appended to CD4 (CD4 ζ 10; Table 1), however, was not retained in the ER, suggesting that the homologous sequence in ζ probably does not contain an ER retention signal and that nonspecific conformational changes are unlikely to be responsible for ER retention of CD4 ϵ 10.

We further characterized the ER retention signal by mutagenesis of the CD3 ϵ tag sequence. Addition of KGQR (CD4mt32) or QRDL (CD4 ϵ 4) to CD4 were not sufficient to promote retention in the ER, but a chimaeric CD4 protein with DLYSGL (CD4 ϵ 6) was partially retained, suggesting that this sequence contained part of the retention signal. Mutation of the serine and tyrosine residues of the tag sequence (CD4mt18 and CD4mt12) but not of the second glycine, lysine and arginine (CD4mt6, CD4mt24, CD4mt34) resulted in loss of ER retention. Tyrosine at position 177 in the ER retention sequence is part of a consensus sequence also found in the C termini of subunits of receptors such as CD3 γ and CD3 δ (ref. 8). Sequences homologous to those of the CD3 ϵ ER retention motif may conceivably be involved in the retention of other multisubunit receptors in the ER. The finding that the arginine residue of the retention motif is expendable is supported by the presence of glutamine in this position in canine CD3 ϵ (ref. 9).

Our results provide evidence for an ER retention signal in the amino-acid sequence spanning residues 171 to 180 of the intracytosolic tail of CD3 ϵ that differs from previously described signals^{10,11}. Tac-CD3 ϵ chimaeras containing this cytoplasmic region of CD3 ϵ can be expressed on the surface of transfected T cells¹², but the surface expression of this chimaera could be due to interactions with other endogenous T-cell receptor (TCR)

or interleukin-2 receptor chains of the transfected cell. Elimination of the transmembrane domain of CD3 ϵ also prevented ER retention, so CD3 ϵ may contain two ER retention sequences, one in the cytoplasmic tail in the region defined by deletion 11, and the other in the transmembrane domain. But deletion of amino acids 171-180 of the cytoplasmic tail is sufficient to abolish ER retention of CD3 ϵ and furthermore, deletion of the transmembrane domain may indirectly prevent ER retention as a result of the location of this anchor-free mutant in the lumen of the ER blocking the interaction of the 171-180 sequence with its putative cytosolic receptor. A model is favoured, therefore, in which only the motif in the cytoplasmic tail of CD3 ϵ is physiologically important in the retention of this polypeptide in the ER.

It has been suggested²⁴ that the presence of charged amino acids in the transmembrane domains of the TCR and CD3 chains is critical for their retention in the ER, their assembly and rapid degradation. Furthermore, the introduction of a charged amino acid in a central position of the transmembrane domain of the Tac protein results in its retention and degradation in the ER¹³. But, there are data that argue against a general role for charged transmembrane residues in such processes: CD3 ϵ has an aspartic acid in position 11 of the transmembrane domain but is stable, and a mutant with the transmembrane aspartic acid substituted for alanine is still retained in the ER¹⁴. Our results show that CD3 ϵ contains an ER retention sequence outside the transmembrane region and that deletion mutant 11, with an intact transmembrane region, is transported to the cell surface.

A model is proposed in which the assembly and expression at the cell surface of the TCR/CD3 complex is a multistep process involving association of CD3 ϵ with either CD3 γ or CD3 δ to form a dimer which interacts with TCR α and TCR β (refs 15, 16). The $\delta\epsilon\alpha\beta$ and $\gamma\epsilon\alpha\beta$ tetramers can then interact

with dimers of ζ or homologues (η and Fc ϵ receptor γ -chain) to form a complete receptor complex. Only with the assembly of an intact complex would all the retention signals be hidden from their putative cognate receptors and allow the surface expression of the TCR/CD3 complex. □

Received 20 January; accepted 13 April 1992.

- Ashwell, J. D. & Klausner, R. D. *A. Rev. Immunol.* **8**, 139–167 (1990).
- Clevers, H. C., Alarcón, B., Wileman, T. E. & Terhorst, C. A. *Rev. Immunol.* **6**, 629–662 (1988).
- Lippincott-Schwartz, J., Bonifacio, J. S., Yuan, L. C. & Klausner, R. D. *Cell* **54**, 209–220 (1988).
- Wileman, T., Carson, G. R., Concino, M., Ahmed, A. & Terhorst, C. *J. Cell Biol.* **110**, 973–986 (1990).
- Terasaki, M., Song, J., Wong, J. R., Weiss, M. J. & Chen, L. B. *Cell* **38**, 101–108 (1984).
- Weissman, A. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 9709–9713 (1988).
- Romeo, C., Amiot, M. & Seed, B. *Cell* **68**, 889–897 (1992).
- Reth, M. *Nature* **338**, 383–384 (1989).
- Nash, R. A., Scherf, U. & Storb, R. *Immunogenetics* **33**, 396–398 (1991).
- Munro, S. & Pelham, H. R. B. *Cell* **48**, 899–907 (1987).
- Nilsson, T., Jackson, M. R. & Peterson, P. A. *Cell* **58**, 707–718 (1989).
- Letourneur, F. & Klausner, R. D. *Science* **255**, 79–82 (1992).
- Bonifacio, J., Cosson, P., Shah, N. & Klausner, R. D. *EMBO J.* **10**, 2783–2793 (1991).
- Hall, C. *et al. Int. Immun.* **3**, 359–368 (1991).
- Alarcón, B. *et al. EMBO J.* **10**, 903–921 (1991).

- Kappes, D. J. & Tonegawa, S. *Proc. natn. Acad. Sci. U.S.A.* **88**, 10619–10623 (1991).
- Inouye, S. & Inouye, M. in *Directed Mutagenesis* (ed. McPherson, M. J.) 71–82 (Oxford University Press, New York, 1991).
- Takebe, Y. *et al. Molec. cell. Biol.* **8**, 466–472 (1988).
- Blumberg, R. S. *et al. Proc. natn. Acad. Sci. U.S.A.* **87**, 7220–7224 (1990).
- Pagano, R. E. *Meth. Cell Biol.* **29**, 75–85 (1989).
- Salmerón, A., Sánchez-Madrid, F., Ursa, M. A., Fresno, M. & Alarcón, B. *J. Immunol.* **147**, 3047–3052 (1991).
- Carrera, A. C., Sánchez-Madrid, F., López-Botet, M., Bernabeu, C. & Landazuri, M. O. *Eur. J. Immunol.* **17**, 179–186 (1987).
- Alarcón, B., Berkhout, B., Breitmeier, J. & Terhorst, C. *J. biol. Chem.* **263**, 2953–2961 (1988).
- Bonifacio, J. S., Cosson, P. & Klausner, R. D. *Cell* **63**, 503–513 (1990).

ACKNOWLEDGEMENTS. We thank J. Avila, P. Bonay, S. Ley and O. Williams for comments and critically reading the manuscript; M. Chorro and M. J. Serramia for technical assistance; and T. Nilsson for the CD8/E3.9 chimera. This work was supported by grants from CICYT, Fundación Ramón Areces, and Comunidad de Madrid.

Antigen presentation enhanced by the alternatively spliced invariant chain gene product p41

Mary Peterson & Jim Miller

Department of Molecular Genetics and Cell Biology, University of Chicago, 920 East 58th Street, Chicago, Illinois 60637, USA

DURING biosynthesis, class II molecules of the major histocompatibility complex are associated with a nonpolymorphic protein called invariant chain, Ii, which facilitates folding of class II molecules and their exit from the endoplasmic reticulum^{1–4}, interferes with their association with peptide^{5,6} and directs their post-Golgi transport (refs 7–9). If Ii blocks class II loading with endogenous antigens in the endoplasmic reticulum and/or directs class II molecules to the exogenous antigen-loading compartment, then the co-expression of Ii should enhance the ability of class II molecules to present exogenous antigens to T cells. But data supporting a role for Ii in class II-restricted antigen presentation are controversial^{1,10–13}. Here we show that Ii can facilitate exogenous antigen presentation for a subset of antigens. Although all known functions of Ii have been ascribed to the principal form of Ii, p31, we find that in most cases antigen presentation is facilitated only by the alternatively spliced, minor form of Ii, p41.

The murine *Ii* gene encodes two polypeptide chains, one (p31) of relative molecular mass 31,000 (31K) and another less abundant (~10% of total Ii) 41K protein (p41) that contains 64 extra amino acids encoded by an alternatively spliced exon^{14,15}. Although both forms of Ii seem to associate equally well with class II molecules, our previous analysis of the function of Ii was limited to the more abundant p31 form¹. To determine whether p41 has a role in class II biosynthesis and antigen presentation, we transfected class II-positive Ltk⁻ cells with *Ii* clones encoding p31 alone, p41 alone, or p31 and p41 together (expressed from a genomic *Ii* clone) (Fig. 1). Analysis of cells transfected with either p31 or p41 confirmed that both forms of Ii can increase the efficiency of class II transport from the endoplasmic reticulum (ER) to the Golgi (data not shown)^{2–4} and induce the conformational effects in I-A^d necessary to generate the Ii-dependent monoclonal antibody epitopes (Fig. 2)¹. These data indicate that both p31 and p41 associate with class II in a way that can modify class II folding and early transport events. In contrast to I-A^d, flow cytometric analysis did not reveal any Ii-dependent conformational effects on the I-A^k molecule (Fig. 2). Although this could reflect inherent

differences in the effect of Ii on the folding of individual alleles of class II, we cannot exclude differences in the specificity or avidity of the I-A^d and I-A^k-reactive antibodies tested.

Although the ability of p31 and p41 to facilitate early events in class II biosynthesis is comparable, the ability to enhance antigen presentation to T cells is mostly limited to p41 (Fig. 3, and Table 1). With some T cells Ii had no effect on the ability of the L-cell transfectants to process and present antigen (Fig. 3a–c), whereas with others, co-expression of Ii could significantly enhance antigen presentation (Fig. 3d–g). These results are consistent with a report that Ii can facilitate antigen

TABLE 1 Ability of Ii to enhance antigen presentation is specific for the peptide MHC complex that is generated and often restricted to the p41 form of Ii

	T cell	Specificity	Ii-dependence		Ref.
			p41	p31	
(a)	3D0-54.8	I-A ^d /OVA _{323–339}	1	1	23, 24
	HMV 29.4	I-A ^d /HMyo	1	1	*
	Pd-2.25.3	I-A ^d /PINS	1	1	25
	h4Ly18.2	I-A ^k /HEL	1	1	26
	TS12	I-A ^k /RNase _{40–61}	1	1	27
(b)	3A9	I-A ^k /HEL _{46–61}	100	10	28
	SKK45.10	I-A ^k /KLH	100	10	29
	SKK9.11	I-A ^k /KLH	10	10	29
(c)	3D0-18.3	I-A ^d /OVA	100	1	23
	C7R14	I-A ^d /SWMMyo _{102–118}	50	1	30
	GD5.16	I-A ^d /HEL _{20–35}	10	1	†
	GD3.25.5	I-A ^d /HEL _{11–25}	10	1	†
	AODH-3.4	I-A ^k /OVA	10	1	31

Data from Fig. 3 and additional data not shown are summarized here. Antigen presentation to some T cells is independent of Ii (a), whereas for others (b, c), the presence of Ii can increase the efficiency of antigen presentation by 10–100× (Ii-dependence is summarized as the magnitude of the shift in the dose response relative to Ii-negative cells). In most cases the effect of Ii is limited to the p41 form (c), although in some cases expression of only p31 can facilitate antigen presentation (b). The MHC/antigen specificity for each T cell is listed and, where known, the specific peptide is shown. Although the specific peptides recognized by h4Ly18.2 and 3D0-18.3 have not been identified, they are distinct from those recognized by 3A9 and 3D0-54.8 (refs 23, 26 and 28). Details of T-cell hybridomas provided by A. Chervonsky (*) and E. Sercarz (†) are unpublished. With some T-cells (h4Ly18.2, SKK9.11 and AODH-3.4) the function of p41 was only assessed by expression from the genomic *Ii* (encoding p31 and p41) clone. Abbreviations: OVA, chicken ovalbumin; HMyo, horse myoglobin; PINS, pork insulin; HEL, hen egg lysozyme; RNase, bovine ribonuclease; KLH, keyhole limpet haemocyanin; SWMMyo, sperm whale myoglobin.

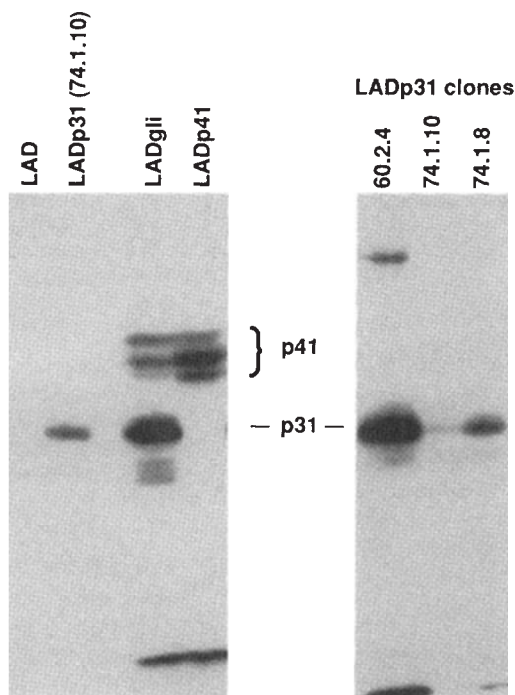


FIG. 1 p31 and p41 li expression in transfected cells lines. Lysates from I-A^d-positive L-cell transfectants expressing no li (LAD, MT72.2.3), p31 li (LADp31, MT74.1.10), both p31 and p41 li (from genomic li) (LADgII, MT73.1.5), or p41 li (LADp41, MT70.2.2) were electrophoresed on 10% SDS-polyacrylamide gels, transferred to nitrocellulose, and invariant chain detected by binding of the li-specific monoclonal antibody, In-1. The blot was developed with mouse-anti-rat immunoglobulin and ¹²⁵I-labelled protein A and autoradiographed; p31 and p41 li are indicated. Several independent p31 transfectants (MT60.2.4, MT74.1.10 and MT74.1.8) that express different amounts of p31 li are shown on the right. Antibody In-1 is specific for the N terminus of li and in addition to intact li, it reacts with N-terminal subfragments generated late in li biosynthesis, presumably by endosomal proteases^{20,32}. These can be seen in cells expressing high levels of li as small bands migrating at the dye front on these gels and occasionally as intermediate fragments of 20–25K. p41 migrates as three distinct bands on the western blots, possibly reflecting differential glycosylation of the two additional N-linked glycosylation sites in p41.

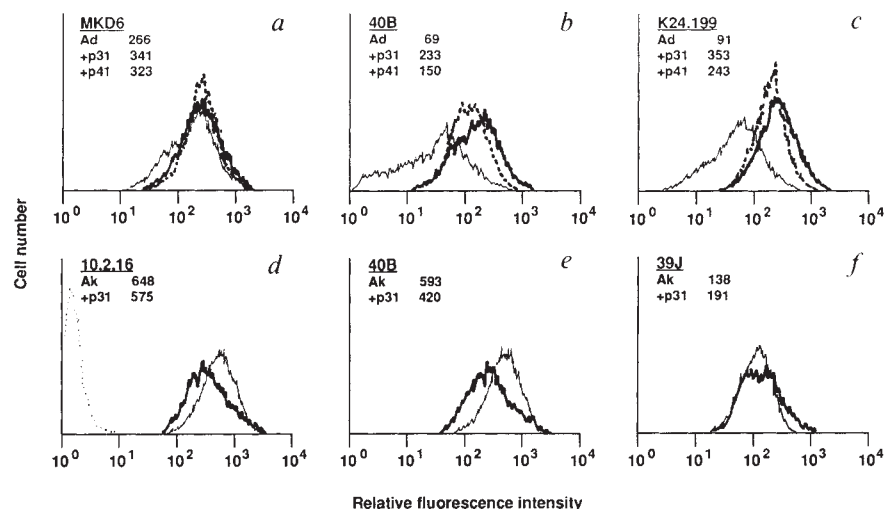
METHODS. Ltk⁻ cells were transfected with cDNA clones encoding I-A^d or I-A^k and class II-positive cells were cloned by limiting dilution as described⁴. Class II-positive cells were then transfected with the gpt-selectable marker only, or with one of three li DNA clones, p31 cDNA, p41 cDNA or genomic (encoding both p31 and p41). This generated cells that were matched for class II expression and varied only for li expression. To avoid artefacts of clonal variation, 3–6 individual clones for each of the 4 genotypes were isolated and analysed for class II expression by flow cytometry (Fig. 2), for li expression by northern and western blotting, and for antigen presentation function (Fig. 3).

presentation in a way that is antigen-specific¹³. The ability of li to enhance antigen presentation showed no correlation with the class II allele or with the antigen, and in some cases it varied for T cells that recognize different peptides from the same protein (see 3A9 and H4Ly18.2, or 3DO-18.3 and 3DO-54.8 in Fig. 3 and Table 1). It did correlate with the expression of p41, however. In one case p31 and p41 both helped antigen presentation (Fig. 3g), but cells expressing only p31 usually presented antigen comparably to or only marginally better than li-negative cells (Fig. 3 and Table 1). This inability of p31 to affect antigen presentation was not due to quantitative differences in li expression because transfectants expressing very different amounts of p31 (Fig. 1) all present antigen to an equivalent extent (Fig. 3h). In addition, complete DNA sequence analysis of the p31 complementary DNA clone confirmed that it encoded a wild-type version of murine p31 (data not shown)¹⁶. Thus, p41 may have a unique role in class II-restricted antigen presentation events.

We found that p41 was able to enhance antigen presentation in cells to the same extent even when its expression varied from ~10% (from the genomic clone) to 100% (from the p41 cDNA clone) of total li (Fig. 3d–f). This indicates that the low level of p41 expressed by alternative splicing of the *li* gene may be saturating for this function. Furthermore, overexpression of p41 did not allow for enhanced antigen presentation to T cells that were not influenced by the presence of the genomic li clone (Fig. 3a, b). Taken together, these data indicate that li can facilitate antigen presentation, but that this facilitation is limited to a subset of antigens, is specific for the peptide-class II complex recognized by individual T cells, and usually requires the expression of the p41 form of li.

We propose the following model for the function of li in class II biosynthesis and function. Both p31 and p41 associate with class II rapidly after biosynthesis, facilitating class II folding in the ER and transport to the Golgi (Fig. 2)^{2–4}. It may function as a surrogate peptide in the ER, interfering with peptide

FIG. 2 Both p31 and p41 forms of li can modify the conformation of class II expressed at the cell surface. a, b and c, Ltk⁻ cells expressing I-A^d alone (MT72.2.7, Ad, thin line), I-A^d plus p31 li (MT74.1.8, +p31, thick line) or I-A^d plus p41 li (MT70.2.4, +p41, dashed line) were stained with class II-specific antibodies, MKD6 (a), 40B (b) or K24-199 (c). Both p31 and p41 induce similar conformational effects in I-A^d. d, e and f, Ltk⁻ cells expressing I-A^k alone (MT64.1.3, Ak, thin line) or I-A^k and p31 li (MT61.1.1, +p31, thick line) were stained with 9 different I-Ak-reactive monoclonal antibodies that differ in their specificity for class II (refs 33, 34): 10.2.16 (d), 40B (e), 39J (f), H116-32, 11-5-2, 26-7-11, 26-8-16, 40F and 40M (not shown). None of these I-A^k antibodies effectively distinguish between li-negative and li-positive cells. The most dramatic differences between li-positive and li-negative cells are with the 40B and 39J antibodies, where the presence of li induced a 1.39 (± 0.22; n=6)-fold increase in 39J staining relative to 40B staining. Note that the cross-reactive 40B antibody is li-dependent on I-A^d, but not on I-A^k. The dotted line in d represents cells stained with the secondary fluorescein isothiocyanate-goat-anti-mouse immunoglobulin reagent only. Values for



mean fluorescence intensity are given in the upper left-hand corner of each panel.

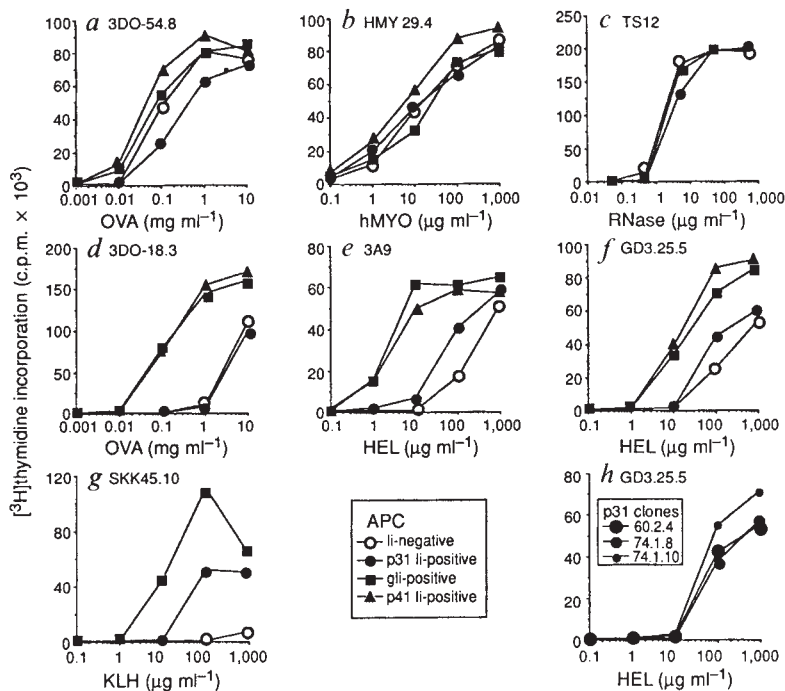


FIG. 3 li-dependent enhancement of antigen presentation is encoded predominantly by p41. Ltk⁻ transfectants expressing I-A^d (a, b, d, f, h) or I-A^k (c, e, g) without li (circles), with p31 li (filled circles), with both p31 and p41 (gli; filled squares), or with p41 li (filled triangles) were assayed for their ability to present various doses of antigen to class II-restricted T-cell hybridomas. See Table 1 for a description of T-cells and their specificities. a-c, li-independent antigen presentation; d-g, li-dependent antigen presentation; h, level of p31 expression does not influence the efficiency of antigen presentation. Data represent the incorporation of $[^3\text{H}]$ thymidine (c.p.m. $\times 10^3$) by the IL-2 dependent T-cell line, CTLL, when cultured in the presence of 25% culture supernatant from the antigen presentation assays. Analysis of multiple independent clones for each genotype gave similar results. Data are presented as the mean for 3-6 independent transfectants for each genotype, except for h and e, for which single representative clones are shown. Transfectants expressing only p41, not shown in c and g, present antigen equivalently to cells expressing genomic li (see Table 1).

loading^{5,6} and enhancing folding. Although this function of Ii may be important for suppressing the presentation of peptides found in the ER, it may also reflect the necessity to protect class II from associating with newly synthesized and partially unfolded proteins in the ER. Nevertheless, the ability of Ii to protect class II from loading with peptides early in biosynthesis does not result in a generalized increase in the efficiency of peptide loading later in biosynthesis (Table 1). The ability of class II-positive transfectants to process and present soluble protein antigens in the absence of Ii (Fig. 3a-c) suggests that class II itself may express the necessary sorting signals for intersection with endocytosed antigen. This implies that the endosomal localization signal in the cytosolic tail of Ii (refs 7, 8) may function by retention, rather than by targeting, of the class II-Ii complex in endosomes. Proteolytic events that generate antigenic peptides and cleave Ii release this retention signal and allow class II to transport to the cell surface (G. Loss and A. Sant, manuscript submitted). This retention function of Ii may account for the ability of p31 to facilitate the presentation of some antigens (Fig. 3g)¹¹.

The unique ability of p41 to enhance antigen presentation argues that p41 has an additional function beyond the ability of Ii to block peptide loading and to retain class II in endosomes. A model for the role of p41 must take into account both the peptide specificity and the apparent nonstoichiometry of p41 function (Table 1 and Fig. 3). Because it is difficult to determine the relative concentrations of p41-associated class II and antigenic peptide in endosomes, we cannot eliminate a model in which p41 imparts a direct effect on class II-peptide association. We favour the possibility, however, that p41 may contain a distinct intracellular transport signal directing class II from an early to a late endosomal compartment. This putative transport signal in p41 may be encoded in sequences homologous to thyroglobulin¹⁵, a protein that after secretion and storage in thyroid follicles is internalized and degraded in lysosomes to release iodotyrosine¹⁷. In this model, p41 would enhance the presentation of peptides that are released only after extensive proteolysis of the native antigen or which require a lower-pH compartment for association with class II (refs 18, 19, 35). These conditions may only be met after internalization of antigen to a highly acidified late endosomal compartment, rich in proteases. In support of this idea, class II has been localized to late endo-

somes^{20,21}. Those peptides whose presentation is not enhanced by the presence of p41 may be generated early after endocytosis and may not require the transport of class II from the early to the late endosome. The apparent nonstoichiometry of p41 function may be accounted for by the formation of multimeric complexes of class II and Ii²², allowing a single p41 molecule within the complex to direct the transport of an entire cohort of class II molecules. □

Received 14 February; accepted 8 May 1992.

- Peterson, M. & Miller, J. *Nature* **345**, 172-174 (1990).
- Layet, C. & Germain, R. *Proc. natn. Acad. Sci. U.S.A.* **88**, 2346-2350 (1991).
- Schaff, W., Hruska, K., Bono, C., Shuman, S. & Schwartz, B. *J. Immun.* **147**, 603-608 (1991).
- Anderson, M. & Miller, J. *Proc. natn. Acad. Sci. U.S.A.* **89**, 2282-2286 (1992).
- Roche, P. & Cresswell, P. *Nature* **345**, 615-618 (1990).
- Teyton, L. *et al.* *Cell* **348**, 39-44 (1990).
- Bakke, O. & Dobberstein, B. *Cell* **63**, 707-716 (1990).
- Lotteau, V. *et al.* *Nature* **348**, 600-605 (1990).
- Lamb, C., Yewdell, J., Bennink, J. & Cresswell, P. *Proc. natn. Acad. Sci. U.S.A.* **88**, 5998-6002 (1991).
- Sekaly, R. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **85**, 1209-1213 (1988).
- Stockinger, B. *et al.* *Cell* **56**, 683-689 (1989).
- Bertolino, P. *et al.* *Int. Immun.* **3**, 435-443 (1991).
- Nadimi, F. *et al.* *Eur. J. Immun.* **21**, 1255-1263 (1991).
- Yamamoto, K., Koch, N., Steinmetz, M. & Hammerling, G. *J. Immun.* **134**, 3461-3476 (1985).
- Koch, N., Lamer, W., Habicht, J. & Dobberstein, B. *EMBO J.* **6**, 1677-1683 (1987).
- Miller, J., Hatch, J., Simonis, S. & Cullen, S. *Proc. natn. Acad. Sci. U.S.A.* **85**, 1369-1363 (1988).
- Ekholm, R. *Int. Rev. Cytol.* **120**, 243-288 (1990).
- Harding, C., Roof, R., Allen, P. & Unanue, E. *Proc. natn. Acad. Sci. U.S.A.* **88**, 2740-2744 (1991).
- Jensen, P. *J. exp. Med.* **171**, 1779-1784 (1990).
- Pieters, J., Horstmann, H., Bakke, O., Griffiths, G. & Lipp, J. *J. Cell Biol.* **115**, 1213-1223 (1991).
- Peters, P., Neefjes, J., Oorschot, V., Ploegh, H. & Geuze, H. *Nature* **349**, 669-676 (1991).
- Roche, P., Marks, M. & Cresswell, P. *Nature* **354**, 392-394 (1991).
- Shimonkevitz, R., Kappler, J., Marrack, P. & Grey, H. *J. exp. Med.* **158**, 303-316 (1983).
- Shimonkevitz, R., Colon, S., Kappler, J., Marrack, P. & Grey, H. *J. Immun.* **133**, 2067- (1984).
- Jensen, P. *J. Immun.* **141**, 2545-2550 (1988).
- Rosoniec, E. *et al.* *J. Immun.* **143**, 50-58 (1989).
- Lorenz, R., Tyler, A. & Allen, P. *J. Immun.* **141**, 4124-4128 (1988).
- Allen, P., Strydom, D. & Unanue, E. *Proc. natn. Acad. Sci. U.S.A.* **81**, 2489-2493 (1984).
- Haskins, K. *et al.* *J. exp. Med.* **160**, 452-471 (1984).
- Ozaki, S., Duram, S., Muegge, K., York-Jolley, J. & Berzofsky, J. *J. Immun.* **141**, 71-78 (1988).
- Kappler, J., White, J., Wegmann, D., Mustain, E. & Marrack, P. *Proc. natn. Acad. Sci. U.S.A.* **79**, 3604-3608 (1982).
- Blum, J. & Cresswell, P. *Proc. natn. Acad. Sci. U.S.A.* **85**, 3975-3979 (1988).
- Braunstein, N., Germain, R., Loney, K. & Berkowitz, N. *J. Immun.* **145**, 1635-1645 (1990).
- Pierres, M., Devaux, C., Dosseto, M. & Marchetto, S. *Immunogenetics* **14**, 481-495 (1981).
- Sette, A. *et al.* *J. Immun.* **148**, 844-851 (1992).

ACKNOWLEDGEMENTS. We thank A. Sant for suggesting class II-encoded transport signals, C. Stebbins for proposing that Ii may block association of class II with unfolded proteins in the ER; M. Anderson, A. Chervonsky, G. Loss, M. Naujokas, A. Sant, K. Swier and L. Zuckerman for discussion and comments on the manuscript; P. Allen, J. Berzofsky, A. Chervonsky, J. Freed, P. Jensen, P. Marrack and E. Sercarz for T-cell hybridomas; N. Koch for the genomic li clone; P. Jones for the p41 cDNA clone; and D. Mathis for the I-A^k cDNA clones. Flow cytometry was done in the University of Chicago Cancer Center Core Facility. This work was supported by grants from the NIH and the Cancer Research Institute (J.M.).

Ca²⁺ release induced by inositol 1,4,5-trisphosphate is a steady-state phenomenon controlled by luminal Ca²⁺ in permeabilized cells

Ludwig Missiaen, Humbert De Smedt, Guy Droogmans & Rik Casteels

Laboratorium voor Fysiologie, KUL Campus Gasthuisberg, Herestraat 49, B-3000 Leuven, Belgium

LOW concentrations of inositol 1,4,5-trisphosphate (InsP₃) evoke a very rapid mobilization of intracellular Ca²⁺ stores in many cell types, which can be followed by a further, much slower efflux¹⁻⁸. Two explanations have been suggested for this biphasic release. The first proposes that the Ca²⁺ stores vary in their sensitivity to InsP₃, and each store releases either its entire contents or nothing^{2,5,7} (all-or-none release); the second proposes instead that the stores are uniformly sensitive to the effects of InsP₃, but that they can release only a fraction of their Ca²⁺ before their sensitivity is somehow attenuated^{6,8-11} (steady-state release). Experiments using purified InsP₃ receptor molecules reconstituted into lipid vesicles have shown heterogeneity of the receptors in their response to InsP₃ under conditions in which the total Ca²⁺ level at both sides of the receptor is held constant⁷. We now report that in permeabilized A7r5 smooth-muscle cells incubated in Ca²⁺-free

medium, the amount of ⁴⁵Ca²⁺ remaining in the stores after the rapid transient phase of release is independent of their initial Ca²⁺ levels, indicating that partially depleted stores are less sensitive to InsP₃. Moreover, if the stores are reloaded with ⁴⁰Ca²⁺ after the first stimulus, reapplication of the same low concentration of InsP₃ will release further ⁴⁵Ca²⁺. This recovery of InsP₃ sensitivity is almost complete. Under these conditions, Ca²⁺ release must thus occur by a steady-state mechanism, in which the decreasing Ca²⁺ content of the stores slows down further release.

Permeabilized A7r5 smooth-muscle cells loaded to equilibrium with ⁴⁵Ca²⁺ slowly lost their ⁴⁵Ca²⁺ during incubation in a Ca²⁺-free solution containing 1 mM EGTA and 2 μM thapsigargin (Fig. 1A, a). A short exposure to a submaximal InsP₃ concentration ([InsP₃]; 3 μM) accelerated this Ca²⁺ loss (Fig. 1A, b). The release was completed within 1 s (refs 3, 6, 8), which is too fast to be resolved by our ⁴⁵Ca²⁺-flux technique.

Prolonged incubation with 3 μM InsP₃ (Fig. 1A, c) released more Ca²⁺ than a 2-min exposure (Fig. 1A, b). This sustained phase of the response does not follow a mono-exponential function, but becomes progressively more shallow as the stores lose Ca²⁺. The [InsP₃] was fairly constant, as only a low cell density was used and fresh medium was added every 2 min. Inositol 2,4,5-trisphosphate (Ins(2,4,5)P₃; 2 μM), which is less easily metabolized¹²⁻¹⁴, induced a similar biphasic release (Fig. 1B). This sustained phase of the response cannot be due to a slow dissociation of Ca²⁺ from binding proteins, because higher [InsP₃] triggers a more complete transient release (Fig. 1A, d). Nor are the two phases of the response a delayed consequence of removing Ca²⁺ from the cytoplasmic side of the InsP₃

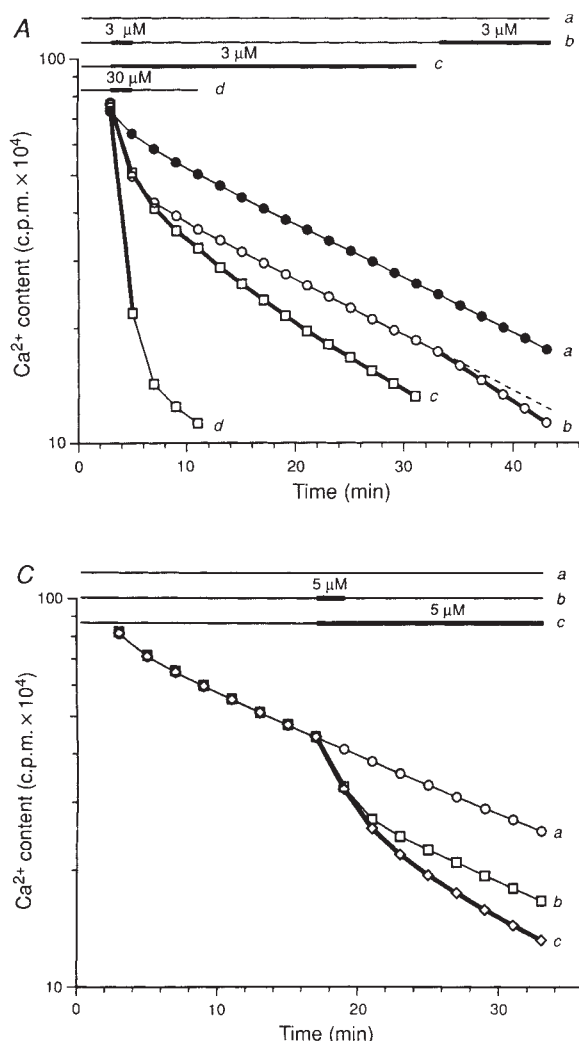


FIG. 1 Continuous Ca²⁺ release after a submaximal InsP₃ stimulation. A, Shows how stimulation with the indicated [Ins(1,4,5)P₃] for the time indicated by the bold line affects the ⁴⁵Ca²⁺ loss from the stores during incubation in Ca²⁺-free solution. B, Compares the effect of a short (b) and long (c) incubation in 2 μM Ins(2,4,5)P₃. C, Compares the effect of a short (b) and prolonged (c) 5 μM Ins(1,4,5)P₃ addition after 17 min of efflux. Each curve is typical of three experiments.

METHODS. ⁴⁵Ca²⁺ fluxes on monolayers of saponin-permeabilized A7r5 cells (3 × 10⁵ cells per 4 cm² well) at 25 °C were measured as described³². The stores were loaded for 10 min in a medium containing 120 mM KCl, 30 mM imidazole (pH 6.8), 5 mM MgCl₂, 5 mM ATP, 0.44 mM EGTA, 10 mM NaN₃ and 150 nM free Ca²⁺ (23 μCi ml⁻¹). The wells were then washed three times in an efflux medium containing 120 mM KCl, 30 mM imidazole (pH 6.8), 2 mM MgCl₂, 1 mM ATP, 1 mM EGTA, 5 mM NaN₃ and 2 μM thapsigargin. This medium (1 ml) was then added to the monolayers and replaced every 2 min. The time course of the ⁴⁵Ca²⁺ wash-out was calculated by summing in retrograde order the amount of ⁴⁵Ca²⁺ remaining in the cells at the end of the experiment and the amounts of ⁴⁵Ca²⁺ collected during the successive time intervals. The efficiency of the washing steps was assessed by measuring the recovery of ¹⁴C-sucrose which was 91.26% after one and 99.04% after two collections. The values for ³H-InsP₃ were 93.70 and 99.26% respectively.

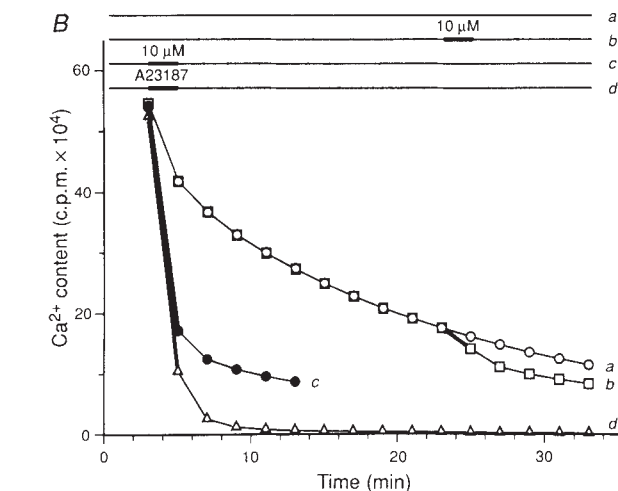
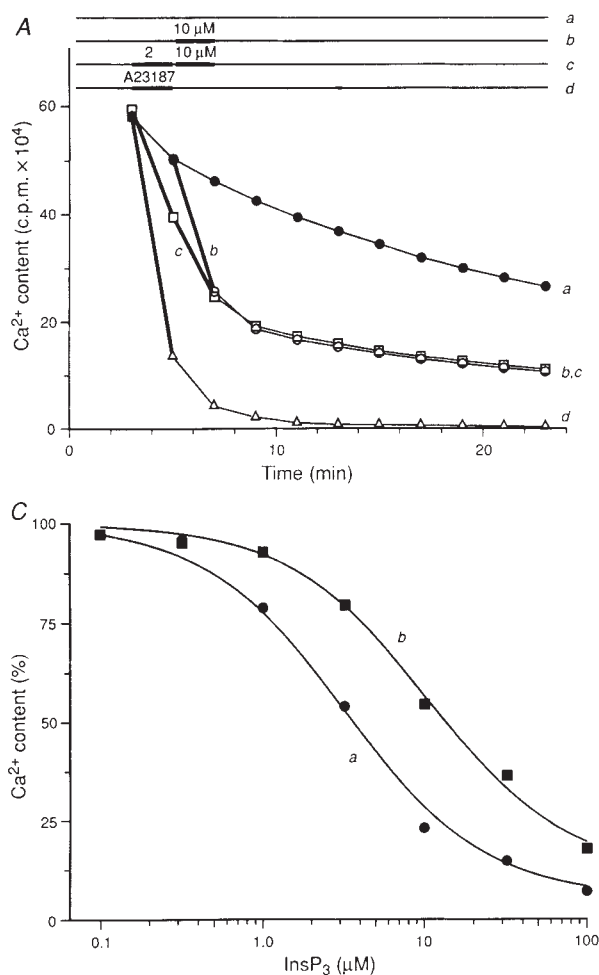


FIG. 2 Luminal Ca^{2+} interferes with the release process. The experimental protocol was the same as in Fig. 1. A, Effects of a 2-min exposure to 10 μM InsP_3 with (c) and without (b) preceding stimulation with 2 μM InsP_3 . B, Effect of a 2-min exposure to 10 μM InsP_3 after 3 min (c) or 23 min (b) of efflux. A, d and B, d, Effect of 10 μM A23187. C, Percentage Ca^{2+} remaining in the stores after a 2-min application of the indicated $[\text{InsP}_3]$ after 3 min (a) or 23 min (b) of efflux. Each curve is typical of three experiments.

receptor, as they are also observed when 5 μM InsP_3 is added after 17 min (Fig. 1C). It has been proposed that Ca^{2+} may be translocated from uptake compartments to structurally different release sites^{15,16}, perhaps through a GTP-dependent mechanism¹⁷. The slow release is not due to such diffusion, because re-addition of 3 μM InsP_3 28 min after the first dose only restores the slow component of the release (Fig. 1A, b). The slow release is also unaffected by 100 μM GTP in the presence of 3% polyethyleneglycol (data not shown). The absence of a second transient release after reapplication of InsP_3 indicates that spontaneous inactivation of the InsP_3 receptor is not responsible for termination of the transient phase of the release^{6,8}, which is consistent with the non-inactivating single-channel recordings during the exposure to InsP_3 for 66 (ref. 18) and 100 s (ref. 19).

Continuous Ca^{2+} release has already been observed by recording the free Ca^{2+} concentration ($[\text{Ca}^{2+}]$) of the medium in permeabilized cells^{3,8}, but the InsP_3 -induced $[\text{Ca}^{2+}]$ rise might have directly interfered with the release process^{6,18,20}. As photo-releasing a small amount of InsP_3 only results in a slow pace-maker $[\text{Ca}^{2+}]$ rise for 5 s in *Xenopus*²¹ and mouse²² oocytes, a continuous release may also occur in intact cells. The slow $^{45}\text{Ca}^{2+}$ release occurring in our EGTA-buffered medium (Fig. 1A) was not previously observed in $^{45}\text{Ca}^{2+}$ -flux experiments^{2,4-6}. Differences in $^{45}\text{Ca}^{2+}$ -flux technique or isoform variability of the InsP_3 receptor²³⁻²⁵ may explain these discrepancies.

We investigated whether luminal Ca^{2+} controls the release process (Fig. 2). The amount of Ca^{2+} remaining in the stores after a 2-min application of 10 μM InsP_3 was similar whether the stores were initially relatively full (Fig. 2A, b) or had been partially depleted by pretreatment with 2 μM InsP_3 (Fig. 2A, c). Similar results were obtained when 10 μM InsP_3 was applied 10 min after washing away the 2- μM addition (data not shown).

This phenomenon is also evident in Fig. 2B, for which the stores were partially depleted by prolonged incubation in efflux medium. After 23 min in efflux medium only 33% of the Ca^{2+} initially in the stores remained, compared with 76% after 3 min (data not shown). Adding 10 μM InsP_3 after 23 min only produced a relatively small release (Fig. 2B, b) and brought the Ca^{2+} content to the same level seen after adding 10 μM InsP_3 to more filled stores (Fig. 2B, c). Stores with a low Ca^{2+} content require higher levels of InsP_3 to deplete them (Fig. 2C, b) than do more filled stores (Fig. 2C, a), confirming earlier findings²⁶.

Figure 3 shows that InsP_3 does not release Ca^{2+} in an all-or-none manner, but rather through a steady-state mechanism which is activated by luminal Ca^{2+} and therefore slows down as the pools become depleted. Stores loaded with $^{45}\text{Ca}^{2+}$ and then stimulated with 5 μM InsP_3 were left unloaded (Fig. 3A, a and B, a) or reloaded with unlabelled Ca^{2+} (Fig. 3A, b, and B, b) and then challenged again with 5 μM InsP_3 . If the release were an all-or-none (complete) emptying of all the pools that can respond to 5 μM InsP_3 ^{2,5,7}, then these pools would only contain $^{40}\text{Ca}^{2+}$ after reloading and would therefore be unable to release $^{45}\text{Ca}^{2+}$ when challenged again with 5 μM InsP_3 . The renewed application of 5 μM InsP_3 , however, released $^{45}\text{Ca}^{2+}$ not released by the first stimulus if the pools were reloaded. This restored response did not occur when 2 μM thapsigargin is included in the reload medium (data not shown). Figure 3C shows that the restoration of the InsP_3 sensitivity is nearly complete after reloading. The addition of 6 μM InsP_3 after 2 min of efflux released 90% of the Ca^{2+} mobilized by 100 μM InsP_3 (Fig. 3C, a, b and c). If the pools were first challenged with 6 μM InsP_3 and reloaded with unlabelled Ca^{2+} , then the second addition of 6 μM InsP_3 again released 74% of the maximal releasable Ca^{2+} (Fig. 3C, d, e, f). When 1 μM InsP_3 was used

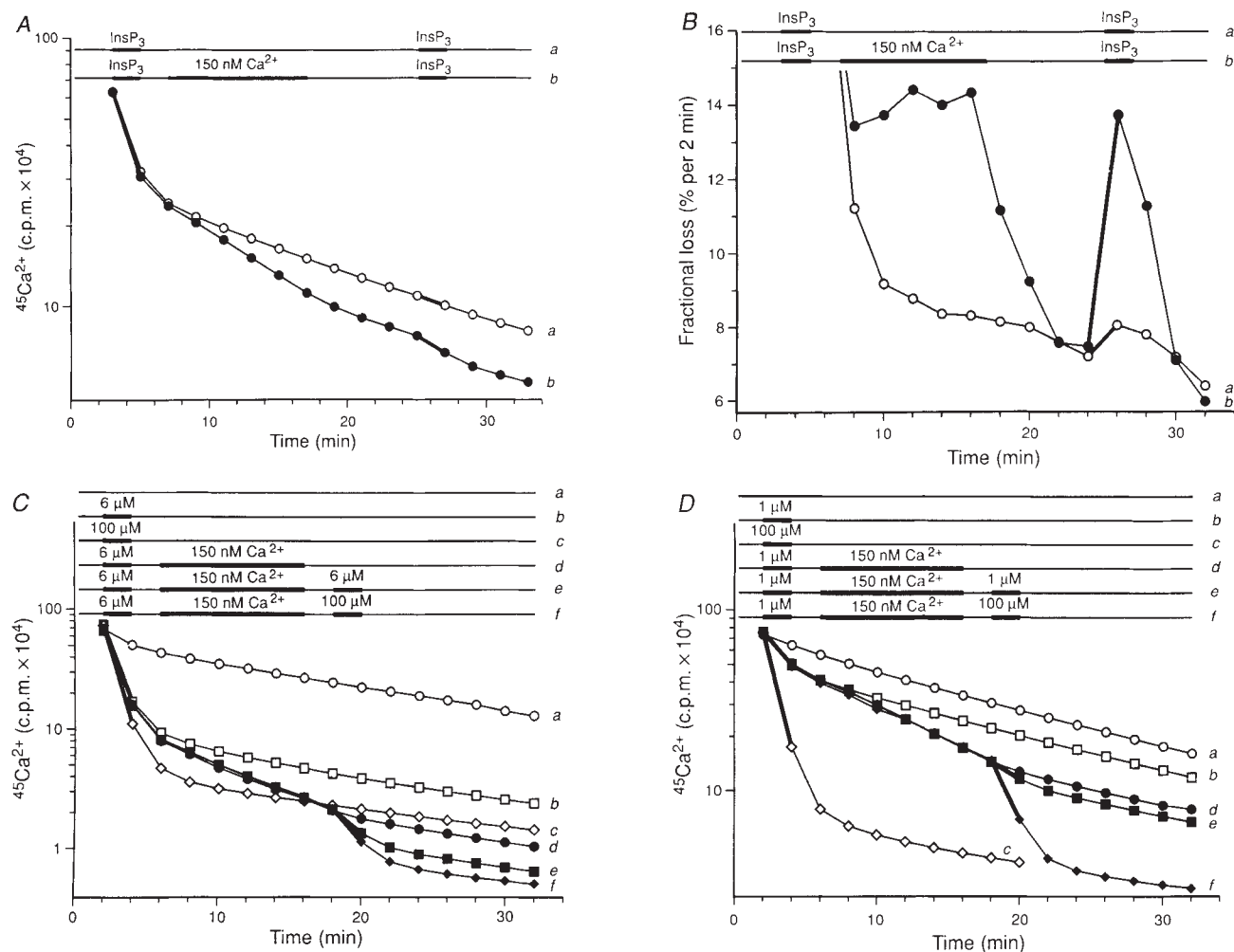


FIG. 3 Reloading the pools with unlabelled Ca^{2+} restores responsiveness to a second InsP_3 application. **A**, **B**, InsP_3 ($5 \mu\text{M}$) was added for 2 min after 3 or 25 min incubation in a medium without thapsigargin. The pools in **b** were reloaded with unlabelled Ca^{2+} after the first InsP_3 addition by increasing the free $[\text{Ca}^{2+}]$ to 150 nM . Note that reloading with unlabelled Ca^{2+} increased the $^{45}\text{Ca}^{2+}$ leak. The results are plotted in **A** as radioactivity versus time

and in **B** as fractional loss (the amount of $^{45}\text{Ca}^{2+}$ leaving the cells relative to the amount of radioactivity in the cells at that time) versus time. The latter parameter is independent of the $[\text{Ca}^{2+}]$ in the stores. **C** and **D**, Effect of stimulation with the indicated $[\text{InsP}_3]$ and reloading with 150 nM unlabelled Ca^{2+} on $^{45}\text{Ca}^{2+}$ loss from the stores. Each curve is typical of three experiments.

instead of $6 \mu\text{M}$, the figures were 32% and 21%, respectively (Fig. 3D). The effects of doses of InsP_3 were also compared with those of $10 \mu\text{M}$ A23187, as luminal Ca^{2+} levels also affect the response to a maximal $[\text{InsP}_3]$ ²⁶. Adding $5 \mu\text{M}$ InsP_3 after 2 min incubation released 71% of the Ca^{2+} released by the ionophore during the next 2 min; if the pools were first depleted by adding $5 \mu\text{M}$ InsP_3 and then reloaded with unlabelled Ca^{2+} , reapplication of $5 \mu\text{M}$ InsP_3 releases 60% of the Ca^{2+} released by the ionophore during the following 2 min (data not shown).

Our data support a model⁹⁻¹¹ in which the Ca^{2+} release mediated by the InsP_3 receptor in the presence of constant cytosolic levels of Ca^{2+} is a steady-state process in which the decreasing luminal levels of Ca^{2+} can limit further release. In this model, the slow phase of the release could be the result of partial channel opening occurring when no more Ca^{2+} is bound to a luminal regulatory site located either on the channel or on some associated luminal protein. The incomplete restoration of InsP_3 sensitivity after reloading (Fig. 3), the pace-maker $[\text{Ca}^{2+}]$ rise culminating in an almost all-or-none Ca^{2+} release in intact cells^{21,27} and the very rapid initial Ca^{2+} release induced by low levels of InsP_3 in permeabilized cells under unbuffered conditions^{3,28} point to a more complete emptying than predicted by the steady-state model. Such emptying is still compatible with the steady-state model if decrease of the luminal $[\text{Ca}^{2+}]$

below the threshold for channel activation only triggered closure of the channel after a time delay or if cytosolic Ca^{2+} exerted positive feedback on the InsP_3 receptor^{6,18,20}. This activation by cytosolic Ca^{2+} can occur even in a Ca^{2+} -buffered medium if the release is faster than the formation of the Ca^{2+} -EGTA complex²⁹. Because the rate of emptying of a Ca^{2+} store will depend on the density of the release channels and on the initial Ca^{2+} content of the stores, the pattern of release (steady-state or all-or-none) may depend not only on the intrinsic properties of the InsP_3 receptor, but also on the density of InsP_3 receptors, Ca^{2+} pumps, luminal Ca^{2+} -binding proteins and passive Ca^{2+} leaks, all of which might differ among various cell types. The intricate kinetics of the release, regulated by both cytosolic^{6,18,20} and luminal Ca^{2+} (refs 26, 30) may be important for setting up complex Ca^{2+} signals³¹ and Ca^{2+} entry^{9,11}. □

Received 4 December 1991; accepted 10 April 1992.

- Berridge, M. J. & Irvine, R. F. *Nature* **341**, 197-205 (1989).
- Muallem, S., Pandolf, S. J. & Beeker, T. G. *J. Biol. Chem.* **264**, 205-212 (1989).
- Meyer, T. & Stryer, L. *Proc. natn. Acad. Sci. U.S.A.* **87**, 3841-3845 (1990).
- Taylor, C. W. & Potter, B. V. L. *Biochem. J.* **266**, 189-194 (1990).
- Oldershaw, K. A., Nunn, D. L. & Taylor, C. W. *Biochem. J.* **278**, 705-708 (1991).
- Finch, E. A., Turner, T. J. & Goldin, S. M. *Science* **252**, 443-446 (1991).
- Ferris, C. D., Cameron, A. M., Hagan, R. L. & Snyder, S. H. *Nature* **356**, 350-352 (1992).
- Champeil, P. et al. *J. Biol. Chem.* **264**, 17665-17673 (1989).

9. Irvine, R. F. *FEBS Lett.* **263**, 5-9 (1990).
10. Tregear, R., Dawson, A. P. & Irvine, R. F. *Proc. R. Soc. B* **243**, 263-268 (1991).
11. Irvine, R. F. *Bioessays* **13**, 419-427 (1991).
12. Morris, A. J., Murray, K. J., England, P. J., Downes, C. P. & Michell, R. H. *Biochem. J.* **251**, 157-163 (1988).
13. Hill, T. D., Dean, N. M. & Boynton, A. L. *Science* **242**, 1176-1178 (1988).
14. Bird, G. St J. *et al.* *Nature* **352**, 162-165 (1991).
15. Menniti, F. S. *et al.* *J. Biol. Chem.* **266**, 13646-13653 (1991).
16. Satoh, T. *et al.* *J. Cell Biol.* **111**, 615-624 (1990).
17. Ghosh, T. K., Mullaney, J. M., Tarazi, F. I. & Gill, D. L. *Nature* **340**, 236-239 (1989).
18. Bezprozvanny, I., Watras, J. & Ehrlich, B. E. *Nature* **351**, 751-754 (1991).
19. Maeda, N. *et al.* *J. Biol. Chem.* **266**, 1109-1116 (1991).
20. Iino, M. *J. gen. Physiol.* **95**, 1103-1122 (1990).
21. Parker, I. & Yao, Y. *Proc. R. Soc. B* **246**, 269-274 (1991).
22. Peres, A. *FEBS Lett.* **275**, 213-216 (1990).
23. Danoff, S. K. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **88**, 2951-2955 (1991).
24. Nakagawa, T., Okano, H., Furuichi, T., Aruga, J. & Mikoshiba, K. *Proc. natn. Acad. Sci. U.S.A.* **88**, 6244-6248 (1991).
25. Südhof, T. C., Newton, C. L., Archer, B. T. III, Ushkaryov, Y. A. & Mignery, G. A. *EMBO J.* **10**, 3199-3206 (1991).
26. Nunn, D. L. & Taylor, C. W. *Molec. Pharmacol.* **41**, 115-119 (1992).
27. Parker, I. & Ivorra, I. *Science* **250**, 977-979 (1990).
28. Meyer, T., Wensel, T. & Stryer, L. *Biochemistry* **29**, 32-37 (1990).
29. Stern, M. D. *Cell Calcium* **13**, 183-192 (1992).
30. Missiaen, L., Taylor, C. W. & Berridge, M. J. *Nature* **352**, 241-244 (1991).
31. Lechleiter, J., Girard, S., Peralta, E. & Clapham, D. *Science* **252**, 123-126 (1991).
32. Missiaen, L. *et al.* *J. Physiol., Lond.* **427**, 171-186 (1990).

Involvement of p21^{ras} in activation of extracellular signal-regulated kinase 2

Alida M. M. de Vries-Smits*,
Boudewijn M. Th. Burgering*, Sally J. Leeverst†,
Christopher J. Marshall† & Johannes L. Bos*

* Laboratory for Physiological Chemistry, University of Utrecht,
Vondellaan 24A, 3521GG Utrecht, The Netherlands

† Chester Beatty Laboratories, Fulham Road,
London SW3 6JB, UK

MANY growth factors upon stimulation of their receptors induce the activity of extracellular signal-regulated kinases, ERKs, also known as MAP kinases^{1,2}. Several of these growth factors also activate the *ras* proto-oncogene product, p21^{ras} (Ras), by stimulating the conversion of the inactive GDP-bound form of Ras to the active GTP-bound form³⁻⁶. We have shown that direct introduction of p21^{ras} oncoprotein into cells in the absence of growth factors activates ERKs within five minutes⁷, which indicates that normal p21^{ras} may be involved in the activation of ERKs by growth factors.

Here we use a recombinant vaccinia virus expressing an interfering mutant of p21^{ras}, Ras^{Asn17}, to investigate this question. In NIH3T3 cells that overexpress the insulin receptor, this recombinant virus inhibits insulin-induced activation of ERK2 completely, but there is no inhibition of insulin-induced activation of phosphatidylinositol-3-kinase. In rat-1 cells the recombinant virus inhibited ERK2 activity induced by platelet-derived growth factor (PDGF) but not by phorbol ester. We conclude that p21^{ras} mediates insulin- and PDGF-induced activation of ERK2.

In cells overexpressing the insulin receptor (A14 cells) insulin induces an increase in the activity of the extracellular signal-regulated kinase, ERK2 (Fig. 1a). ERK2 is a protein kinase of relative molecular mass 42,000 (42K), also called p42MAP2 kinase, which is activated by phosphorylation of threonine and tyrosine residues^{2,8}. Changes in ERK2 activity can be assayed *in vitro* using myelin basic protein as substrate or from the presence of a form of ERK2 with reduced mobility in electrophoresis. This shifted form of ERK2 corresponds to the activated form resulting from threonine and tyrosine phosphorylation⁷. Activation of ERK2 by insulin in A14 cells occurs within five minutes. Activation of protein kinase C (PKC) by the phorbol ester TPA (12-*O*-tetradecanoylphorbol-13-acetate) also leads to activation of ERK2, but insulin-induced activation of ERK2 is not mediated by a TPA-sensitive PKC because

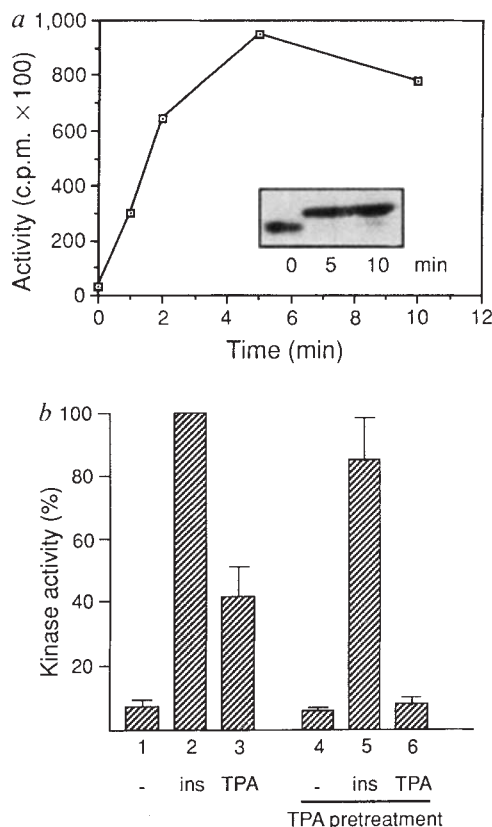


FIG. 1 Insulin-induced activation of ERK2. *a*, Time course of insulin-induced ERK2 activity in A14 cells. Cells were treated with 1 μ M insulin for the indicated time and ERK2 was immunoprecipitated. ERK2 activity in the immunoprecipitates was measured after addition of myelin basic protein as a substrate. The insert shows the mobility shift of the 42K ERK2 protein on stimulation with insulin, as detected by western blotting of total cell lysate. The slower-migrating form of ERK2 corresponds to the activated form. *b*, Effect of TPA-pretreatment on insulin-induced activation of ERK2. ERK2 activity was determined in cells treated with 1 μ M insulin (ins; bars 2 and 5), 100 ng ml⁻¹ TPA (bars 3 and 6) for 10 min, or untreated (bars 1 and 4). For PKC depletion, cells were treated with 100 ng ml⁻¹ TPA for 24 h (bars 4-6) ($n=3$).

METHODS. Cells were starved for 24 h in 0.5% fetal calf serum and treated with insulin or TPA as indicated. Cells were lysed and 100 μ g protein was immunoprecipitated with antiserum 122 as described⁷. The immune complexes were collected using protein A-Sepharose beads, washed and incubated in 30 μ l with 0.4 μ Ci [γ -³²P]ATP in 30 mM Tris, pH 8.0, 20 mM MgCl₂, 2 mM MnCl₂, 10 μ M ATP for 30 min using 7.5 μ g myelin basic protein as substrate. The mixture was electrophoresed on a 15% polyacrylamide gel, blotted onto nitrocellulose and autoradiographed. The labelled protein was counted by scintillation counting. To measure the mobility shift of the ERK2 protein, 50 μ g total lysate was electrophoresed on a 10% polyacrylamide gel, blotted onto nitrocellulose and incubated with antiserum 124 (ref. 7). Immune complexes were detected by horseradish peroxidase-second antibodies, followed by enhanced chemiluminescent detection (Amersham). Protein kinase C was depleted by prolonged treatment with 100 ng ml⁻¹ TPA. For A14 cells, this results in the abolition of TPA-inducible phosphorylation of the 80K MARCKS protein¹⁹ and TPA-inducible activation of ERK2 (see *b*, bars 3 and 6) and the loss of detectable PKC protein in a western blotting experiment using monoclonal anti-PKC (clone MC5; Amersham).

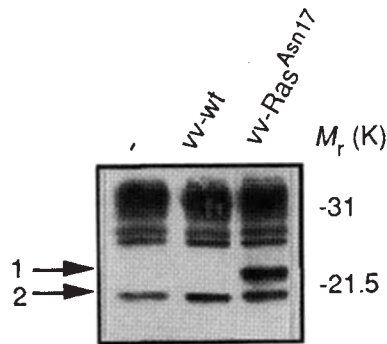


FIG. 2 Expression of vv-Ras^{Asn17} recombinant vaccinia virus in A14 cells. Cell lysate was immunoprecipitated with anti-Ras monoclonal antibody Y13-238, electrophoresed, western blotted and probed with anti-Ras monoclonal Y13-259. Owing to the presence of six histidine residues at the N terminus, Ras^{Asn17} (band 1) has a slightly slower mobility than endogenous p21^{ras} (band 2). *M_r* calibration is shown on the right.

METHODS. Ras^{Asn17} was introduced into the multiple cloning site of plasmid pAgpt containing the consensus sequence for early vaccinia promoters²⁰ (C. Bucci *et al.*, manuscript submitted). The *ras* gene is a human Ha-*ras* gene with 6 additional histidine codons after the AUG start codon and an altered codon 17, where an asparagine codon substitutes a serine codon. The authenticity of the plasmid was confirmed by sequence analysis. Recombination of the plasmid with wild-type vaccinia virus has been described²¹. Recombinant viruses were grown in HeLa cells (titre: 10⁹ virus particles per ml). A14 cells were infected with 10 PFU recombinant or wild-type virus in serum-free medium. After 60 min the medium was replaced and cells were kept in 0.5% fetal calf serum and 10 mM hydroxyurea to block DNA synthesis and late viral gene expression. After 16 h the cells were lysed and 200 µg lysate was immunoprecipitated with Y13-238. The precipitate was separated on a 15% polyacrylamide gel, blotted onto nitrocellulose and incubated with Y13-259. Immune complexes were detected by horseradish peroxidase second antibodies, followed by enhanced chemiluminescent detection.

prolonged pretreatment of the cells with TPA to downregulate PKC does not affect insulin-induced ERK2 activity (Fig. 1b).

We have shown previously that in A14 cells insulin rapidly leads to an increase in the proportion of p21^{ras} in the GTP-bound form³. To investigate whether insulin-induced activation of ERK2 is mediated by p21^{ras}, we constructed a recombinant vaccinia virus expressing a mutant p21^{ras} with asparagine substituted for serine at position 17 (vv-Ras^{Asn17}). Overexpression of Ras^{Asn17} by only a fewfold is sufficient to block p21^{ras} signalling completely, and its introduction into cells is presumed to interfere with the conversion of Ras-GDP to Ras-GTP, perhaps by 'tying up' a GDP/GTP exchange factor involved in the activation of normal p21^{ras} (ref. 9). Although the inactivation of such an exchange factor might affect signalling by other G proteins, we consider that Ras^{Asn17} is likely to be exerting its effect by blocking the activation of normal p21^{ras}, because several effects of expressing Ras^{Asn17} are mimicked by microinjection

of neutralizing antibodies against p21^{ras} (ref. 10) and can be reversed by oncogenic mutant p21^{ras} (ref. 11). A14 cells were infected with wild-type vaccinia virus (vv-wt) and vv-Ras^{Asn17}. After 16 hours in the presence of 10 mM hydroxyurea to block expression of late viral genes, the expression of the introduced Ras^{Asn17} gene was measured (Fig. 2) and found to be present in a 2–3-fold excess over endogenous p21^{ras}. Next we measured the effect of insulin on ERK2 activity in the presence of vv-wt and vv-Ras^{Asn17} (Fig. 3a and b). Infection with vv-wt raised the basal level of ERK2 activity, presumably as a result of the production of EGF-like vaccinia virus growth factor by the virus. Insulin treatment, however, led to a further 1.5–2-fold increase in ERK2 activity and an increased proportion of ERK2 shifted to the slower-migrating form. Infection with vv-Ras^{Asn17} completely inhibited insulin-induced activation of ERK2, resulting in an activity similar to that in unstimulated uninfected cells. These results suggest that p21^{ras} mediates insulin-induced

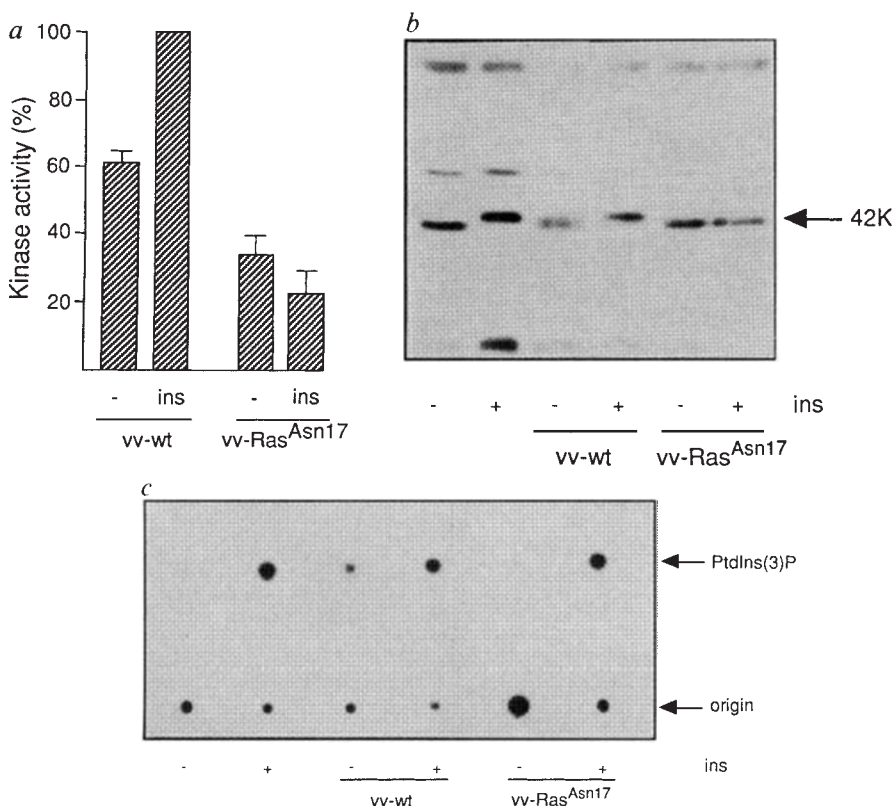


FIG. 3 Ras^{Asn17} recombinant vaccinia virus inhibits insulin-induced activation of ERK2. A14 cells were infected with vv-wt and vv-Ras^{Asn17}, incubated with 1 µM insulin (ins) for 10 min and lysed. **a**, ERK2 activity using myelin basic protein as substrate (*n* = 3). **b**, Mobility shift of ERK2. **c**, PtdIns-3-kinase activity: an autoradiograph of a TLC plate of ³²P-labelled phosphatidylinositol 3-phosphate (PtdIns(3)P) is shown. Infection with vaccinia virus is described in the legend to Fig. 2 and the determination of ERK2 activity and the mobility shift of the 42K ERK protein is described in the legend to Fig. 1. PtdIns-3-kinase activity was determined as described³.

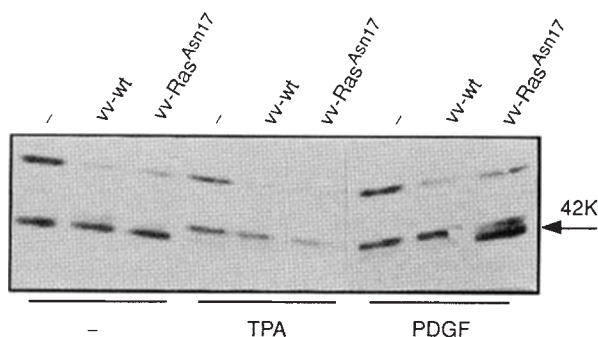


FIG. 4 Ras^{Asn17} recombinant vaccinia virus inhibits PDGF-induced activation of ERK2, but not TPA-induced activation of ERK2. Rat-1 cells were infected with vv-wt or vv-Ras^{Asn17}, incubated with PDGF (5 ng ml⁻¹) or TPA (100 ng ml⁻¹) for 10 min and lysed. Cells were infected as described in the legend to Fig. 2 and the mobility shift determined as described in the legend to Fig. 1.

activation of ERK2. As a control for the effect of Ras^{Asn17} on other parameters of insulin signalling, we measured insulin-induced phosphatidylinositol-3-kinase (PtdIns-3-kinase) activity in the presence of vv-wt and vv-Ras^{Asn17} (Fig. 3c). Tyrosine-phosphorylated PtdIns-3-kinase was collected using an antiserum against phosphotyrosine and its activity measured in the immunoprecipitates using phosphatidylinositol as a substrate. Infection with both vv-wt and vv-Ras^{Asn17} did not affect insulin-induced PtdIns-3-kinase activity, indicating that p21^{ras} does not mediate this part of the insulin signal transduction pathway, a conclusion that has been reached indirectly before³. In cells overexpressing the receptor for insulin-like growth factor 1 (IGF1) and oncogenically activated p21^{ras}, an association between p21^{ras} and IGF1-induced PtdIns-3-kinase has been observed¹². Our results suggest that an interaction between p21^{ras} and PtdIns-3-kinase is not likely to involve p21^{ras} in the regulation of the activity of this kinase, but still leaves open the possibility that the kinase may somehow regulate p21^{ras}. However, in A14 cells we have never detected insulin-induced PtdIns-3-kinase activity in immunoprecipitates of p21^{ras} using monoclonal antibodies Y13-259 or Y13-238 (B.M.Th.B. and J.L.B., manuscript in preparation).

To determine whether the inhibition of ERK2 activity by vv-Ras^{Asn17} is restricted to insulin signalling, we investigated the effect of vv-Ras^{Asn17} on the induction of ERK2 activity by PDGF and TPA in rat-1 cells. As shown in Fig. 4, vv-Ras^{Asn17} inhibits the PDGF-induced ERK2 mobility shift, but not the mobility shift induced by TPA. We conclude that in rat-1 cells p21^{ras} mediates PDGF-induced activation of ERK2 but not TPA-induced activation of ERK2. Also, infection with vv-Ras^{Asn17} does not inhibit TPA-induced activation of ERK2 in A14 cells (not shown). It should be noted that although the activation of ERK2 by insulin and PDGF is completely dependent on p21^{ras}, there is a striking difference in their respective capacities to stimulate an increase in Ras-GTP. Whereas insulin stimulation of A14 cells leads to p21^{ras} being 70% in the GTP-bound form, PDGF has only a small effect on the level of GTP bound to p21^{ras} (refs 3, 5, 6). This may indicate that a very small increase in Ras-GTP, perhaps in conjunction with other PDGF-induced signals, is sufficient for the activation of ERK2.

In Swiss 3T3 cells activation of ERK2 by oncogenic p21^{ras} is channelled through a PKC-dependent and a PKC-independent pathway, although in these cells the PKC-dependent pathway seems to make a greater contribution to ERK2 activation⁷. These results suggest that PKC is downstream of oncogenic p21^{ras}, which is consistent with our observations that Ras^{Asn17} does not block TPA-induced activation of ERK2 in rat-1 cells. But in A14 cells, downregulation of PKC by prolonged exposure to TPA does not block insulin-induced activation of ERK2, showing that p21^{ras} is not mediating an insulin-induced stimulation of PKC. This finding is compatible with results showing that insulin does not activate PKC in NIH3T3 cells overexpressing the insulin receptor¹³. The apparently contradictory results that insulin does not activate PKC and that oncogenic p21^{ras} does can be explained by differences in effect of oncogenically

mutated p21^{ras}, which is constitutively active, and normal p21^{ras}, whose activation is controlled by transmembrane receptors. Oncogenic p21^{ras} may produce global activation of many signal transduction pathways (see ref. 14), whereas with receptors that require normal p21^{ras} it is the receptor rather than p21^{ras} that determines whether a particular signal transduction element such as PKC is activated.

It is becoming evident that the 42-44K ERKs are activated by nerve growth factor through an ERK kinase and a kinase that acts on this kinase¹⁵. For the activation of ERK2 by insulin such a cascade may also occur². It is therefore unlikely that p21^{ras} directly interacts with ERK2, but it might be involved more upstream in the activation cascade of ERK2. One role of activated ERKs appears to be to link cytoplasmic signalling events to nuclear events, because *in vitro* ERK2 phosphorylates the transcriptional activation domain of c-jun at sites also phosphorylated in response to oncogenic p21^{ras} (refs 17, 18). Another function of ERKs is to activate S6 kinase II (ref. 22), and maybe an insulin-sensitive protein kinase that, at least in muscle cells, is involved in the further activation of phosphatase 1G, resulting in the stimulation of glycogen synthesis and the inactivation of glycogenolysis¹⁶. This would link the activation of p21^{ras} to metabolic pathways involved in glucose homeostasis. Although p21^{ras} is not directly involved in insulin-regulated glucose uptake²³, and for muscle cells it has to be shown whether p21^{ras} mediates insulin-induced activation of ERK2, from our results it is clear that p21^{ras} is emerging as an important mediator of insulin signalling. □

Received 17 January; accepted 6 May 1992.

1. Ray, L. B. & Sturgill, T. W. *Proc. natn. Acad. Sci. U.S.A.* **84**, 1502-1506 (1987).
2. Cobb, M. H., Boulton, T. G. & Robbins, D. J. *Cell Reg.* **2**, 965-978 (1991).
3. Burgering, B. M. T. et al. *EMBO J.* **10**, 1103-1109 (1991).
4. Downward, J., Graves, J. D., Warne, P. H., Rayter, S. & Cantrell, D. A. *Nature* **346**, 719-723 (1990).
5. Gibbs, J. B., Marshall, M. S., Scolnick, E. M., Dixon, R. A. F. & Vogel, U. S. *J. biol. Chem.* **265**, 20437-20442 (1990).
6. Satoh, T., Endo, M., Nakafuku, M., Nakamura, S. & Kaziro, Y. *Proc. natn. Acad. Sci. U.S.A.* **87**, 5993-5997 (1990).
7. Levers, S. J. & Marshall, C. J. *EMBO J.* **11**, 569-574 (1992).
8. Anderson, N. G., Maller, J. L., Tonks, N. K. & Sturgill, T. W. *Nature* **343**, 651-653 (1990).
9. Farnsworth, C. L. & Feig, L. A. *Molec. cell. Biol.* **11**, 4822-4829 (1991).
10. Stacey, D. W. et al. *Oncogene* **6**, 2297-2304 (1991).
11. Medema, R. H., Wubbolts, R. & Bos, J. L. *Molec. cell. Biol.* **11**, 5963-5967 (1991).
12. Sjölander, A., Yamamoto, K., Huber, B. E. & Lapetina, E. G. *Proc. natn. Acad. Sci. U.S.A.* **88**, 7908-7912 (1991).
13. Blackshear, P. J., Haupt, D. M. & Stumpo, D. J. *J. biol. Chem.* **266**, 10946-10952 (1991).
14. Cantley, L. C. et al. *Cell* **64**, 281-302 (1991).
15. Gómez, N. & Cohen, P. *Nature* **353**, 170-173 (1991).
16. Dent, P. et al. *Nature* **348**, 302-308 (1990).
17. Pulverer, B. J., Kryakis, J. M., Avruch, J., Nikolakaki, E. & Woodgett, J. R. *Nature* **353**, 670-674 (1991).
18. Binétruy, B., Smeal, T. & Karin, M. *Nature* **351**, 122-127 (1991).
19. Medema, R. H., Burgering, B. M. T. & Bos, J. L. *J. biol. Chem.* **266**, 21186-21189 (1991).
20. Davison, A. J. & Moss, B. *J. molec. Biol.* **210**, 749-769 (1989).
21. Mackett, M., Smith, G. L. & Moss, B. in *DNA Cloning: A Practical Approach* (II) (ed. Glover, D. M.) 191-211 (IRL, Oxford, 1985).
22. Sturgill, T. W., Ray, L. B., Erikson, E. & Maller, J. *Nature* **334**, 715-718 (1988).
23. Osterop, A. P. R. M. et al. *J. biol. Chem.* (in the press).

ACKNOWLEDGEMENTS. We thank M. Zerial for sharing data before publication and for discussion; H. Stunnenberg for the vaccinia virus early promoter expression vector; P. C. van der Vliet and C. P. Verrijzer for vaccinia virus facilities and for help in their preparation; R. Schiphof for HeLa cells; and R. Medema and G. J. Pronk for support and discussions. This work was supported by grants from the Dutch Cancer Society (J.L.B.) and the Cancer Research Campaign (C.J.M.).

The mechanism of Z α_1 -antitrypsin accumulation in the liver

David A. Lomas, Dyfed Ll. Evans, John T. Finch* & Robin W. Carrell

Department of Haematology, University of Cambridge, and
*MRC Laboratory of Molecular Biology, MRC Centre, Hills Road,
Cambridge CB2 2QH, UK

MOST northern Europeans have only the normal M form of the plasma protease inhibitor α_1 -antitrypsin, but some 4% are heterozygotes for the Z deficiency variant¹. For reasons that have not been well-understood, the Z mutation results in a blockage in the final stage of processing of antitrypsin in the liver² such that in the Z homozygote only 15% of the protein is secreted into the plasma. The 85% of the α_1 -antitrypsin that is not secreted accumulates in the endoplasmic reticulum of the hepatocyte; much of it is degraded but the remainder aggregates to form insoluble intracellular inclusions. These inclusions are associated with hepatocellular damage, and 10% of newborn Z homozygotes develop liver disease which often leads to a fatal childhood cirrhosis. Here we demonstrate the molecular pathology underlying this accumulation and describe how the Z mutation in antitrypsin results in a unique molecular interaction between the reactive centre loop of one molecule and the gap in the A-sheet of another. This loop-sheet polymerization of Z antitrypsin occurs spontaneously at 37 °C and is completely blocked by the insertion of a specific peptide into the A-sheet of the antitrypsin molecule. Z

antitrypsin polymerized *in vitro* has identical properties and ultra-structure to the inclusions isolated from hepatocytes of a Z homozygote. The concentration and temperature dependence of this loop-sheet polymerization has implications for the management of the liver disease of the newborn Z homozygote.

The mutation in Z antitrypsin (glutamic acid at position 342 substituted by lysine) involves an amino acid at the base of the reactive centre loop of the molecule³. Using normal M antitrypsin, we have shown that this loop can move in and out of the large A-sheet that forms the major feature of the molecule⁴. This movement of the loop can be induced by mild denaturing conditions and under these same conditions, but at higher temperatures, polymerization can occur between M antitrypsin molecules⁵ (Fig. 1a) owing to the loop of one molecule being inserted into the A-sheet of another (Fig. 2). This polymerization was prevented by the addition of an excess of a synthetic free peptide BC13, which preferentially inserts into the A-sheet forming a binary complex⁶ and so blocks the entry of the reactive centre loop of other antitrypsin molecules. BC13 is a 13-residue peptide derived from the reactive centre loop of antithrombin and homologous to the 345–357 loop sequence of antitrypsin⁴. It contains the highly conserved 345–350 sequence involved in reinsertion of the loop into the A-sheet and forms binary complexes as readily with antitrypsin⁵ as antithrombin ($k_{\text{obs}} = 8 \times 10^{-2} \text{ M}^{-1} \text{ s}^{-1}$). The formation of the binary complex with antitrypsin was confirmed by a concomitant loss of inhibitory activity, a precise change in electrophoretic mobility and a characteristic series of changes in physical properties, including circular dichroic spectra⁴.

In the course of modelling the loop-sheet interaction it

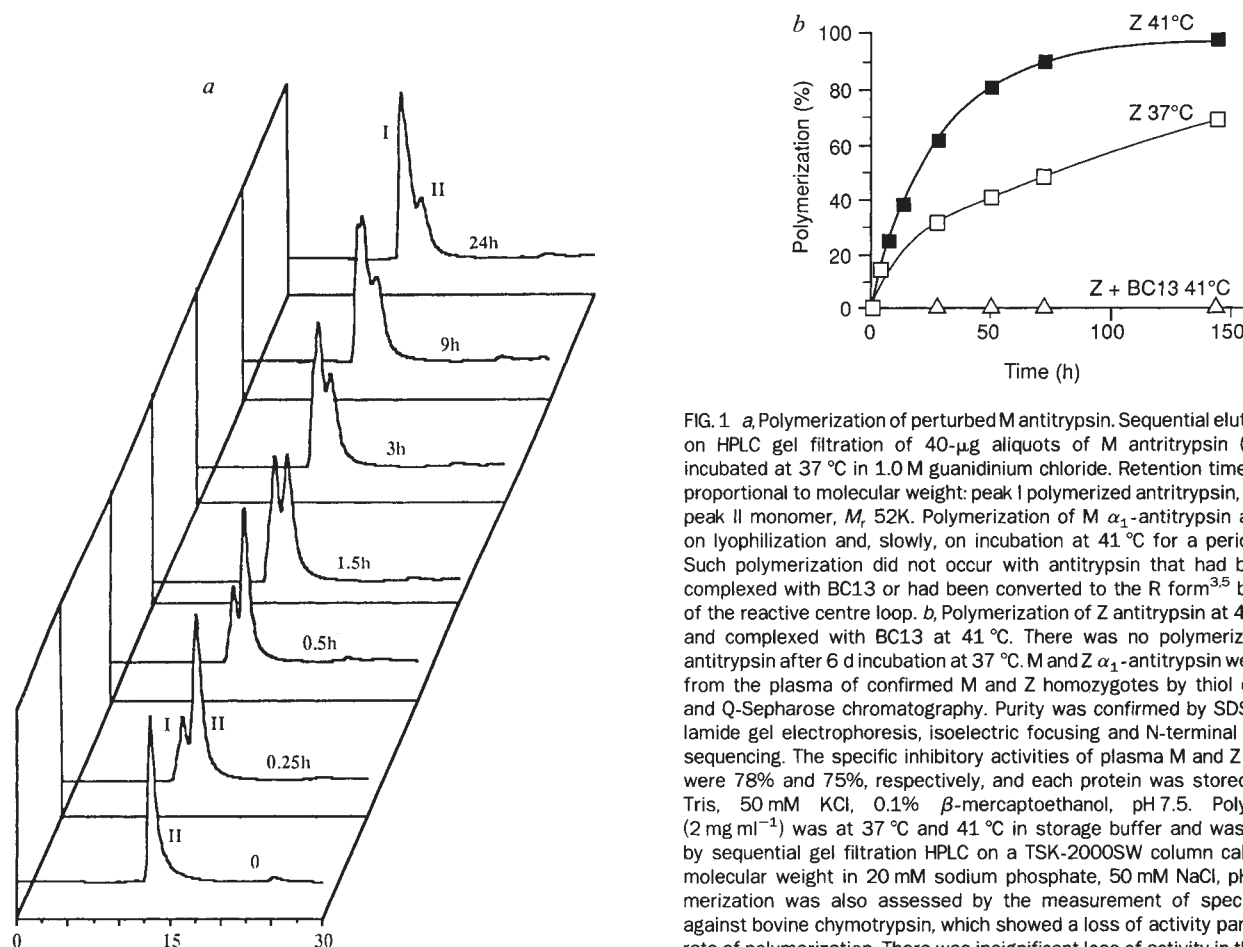


FIG. 1 a, Polymerization of perturbed M antitrypsin. Sequential elution profiles on HPLC gel filtration of 40- μg aliquots of M antitrypsin (2 mg ml^{-1}) incubated at 37 °C in 1.0 M guanidinium chloride. Retention time (x axis) is proportional to molecular weight: peak I polymerized antitrypsin, $M_r > 200\text{K}$; peak II monomer, $M_r 52\text{K}$. Polymerization of M α_1 -antitrypsin also occurs on lyophilization and, slowly, on incubation at 41 °C for a period of days. Such polymerization did not occur with antitrypsin that had been binary complexed with BC13 or had been converted to the R form^{3,5} by cleavage of the reactive centre loop. b, Polymerization of Z antitrypsin at 41 °C, 37 °C, and complexed with BC13 at 41 °C. There was no polymerization of M antitrypsin after 6 d incubation at 37 °C. M and Z α_1 -antitrypsin were isolated from the plasma of confirmed M and Z homozygotes by thiol exchange¹⁵ and Q-Sepharose chromatography. Purity was confirmed by SDS-polyacrylamide gel electrophoresis, isoelectric focusing and N-terminal amino-acid sequencing. The specific inhibitory activities of plasma M and Z antitrypsin were 78% and 75%, respectively, and each protein was stored in 50 mM Tris, 50 mM KCl, 0.1% β -mercaptoethanol, pH 7.5. Polymerization (2 mg ml^{-1}) was at 37 °C and 41 °C in storage buffer and was measured by sequential gel filtration HPLC on a TSK-2000SW column calibrated for molecular weight in 20 mM sodium phosphate, 50 mM NaCl, pH 7.6. Polymerization was also assessed by the measurement of specific activity against bovine chymotrypsin, which showed a loss of activity paralleling the rate of polymerization. There was insignificant loss of activity in the absence of polymerization, even after 12 days incubation at 37 °C. The binary complex⁶ was formed by incubating the antitrypsin (1 mg ml^{-1}) in the storage buffer with a 100-fold molar excess of the loop peptide⁴ at 37 °C for 24 h.

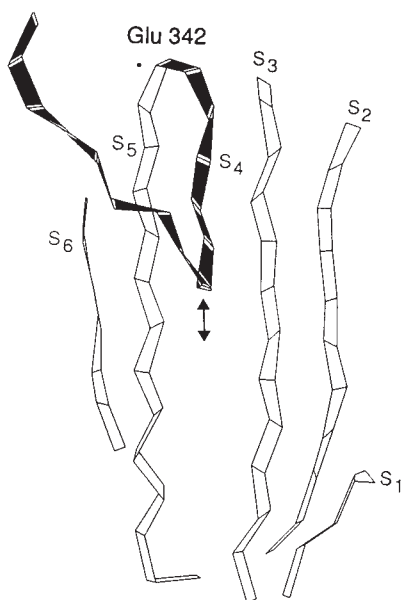
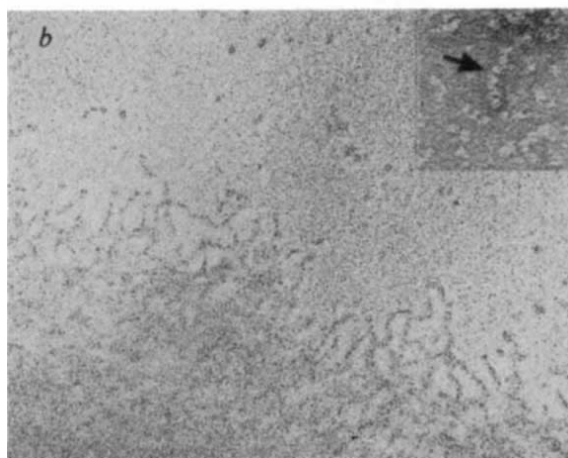
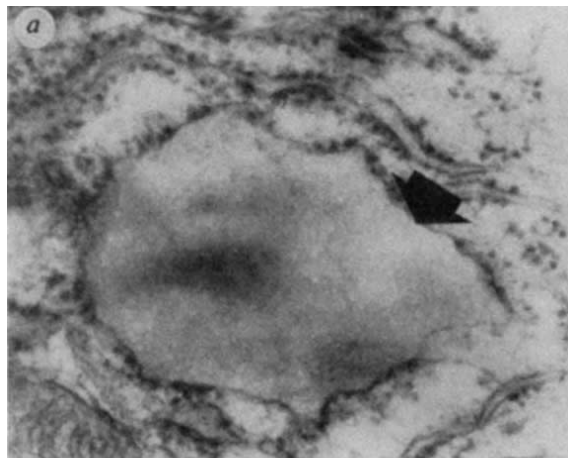


FIG. 2 Loop-sheet polymerization. A diagrammatic representation of the A-sheet of α_1 -antitrypsin. The reactive centre loop (shaded) is mobile and can hinge on Glu 342 to fold back into a gap in the A-sheet between strands 3 and 5 to form strand 4 as shown. Disruption of this folding by the Z mutation Glu 342→Lys or opening of the A-sheet by mild denaturation allows instead the insertion of the loop of a second molecule to give loop-sheet polymerization. This polymerization is prevented by an excess of the free BC13 peptide with a sequence homologous to the loop (shaded), which preferentially inserts into the S_4 position (see ref. 4 for further illustration).



became apparent that the Z mutation at the hinge of the loop could enhance such polymerization. This was confirmed on incubation of isolated M and Z antitrypsin under physiological conditions as shown in Fig. 1b. Z antitrypsin spontaneously polymerized at 37 °C, whereas there was no polymerization of M antitrypsin. Polymerization of Z antitrypsin was greatly accelerated at 41 °C, but the formation of the Z antitrypsin-BC13 complex completely blocks polymerization even after incubation at 41 °C for 6 days. As well as being temperature-dependent, the rate of polymerization of Z antitrypsin was also dependent on concentration, with 30% of 2 mg ml⁻¹ Z antitrypsin being polymerized after 24 h at 37 °C, compared with only 8% at 1 mg ml⁻¹.

This finding of spontaneous loop-sheet polymerization fits with the molecular pathology of the Z antitrypsin. The replacement of the hinge residue 342 (Fig. 2) could either favour extension of the reactive centre loop to make the molecule a donor for dimerization or could allow the A-sheet to open to make the molecule a receptor for dimerization. The circular dichroism spectrum in the near ultraviolet of Z antitrypsin (not shown) supports the second proposal and this also fits well with two other rare antitrypsin variants, M Malton⁷ and Siiyama⁸, that give similar polymerization and hepatocyte accumulation to that seen with the Z variant. Both of these variants have gross mutations at positions 52/53 which will cause a displacement of the B helix³ that forms the base plate for the opening and closing of the A-sheet⁹. Hence the common molecular pathology is likely to be a fixing of the A-sheet in the open position, allowing dimerization with subsequent polymerization.

The loop-sheet mechanism of polymerization provides a satisfying explanation for the hepatocyte inclusions observed in

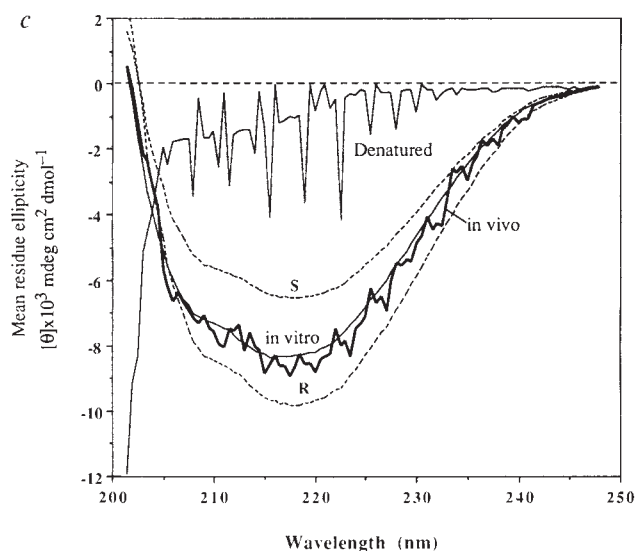


FIG. 3 Liver inclusion. *a*, Electron microscopy ($\times 20,000$) of a hepatocyte from a Z homozygote showing a massive inclusion (large arrow) in the endoplasmic reticulum, and *b*, ($\times 220,000$) showing the filaments of Z antitrypsin protruding from the edge of a hepatic inclusion isolated¹⁰ from the liver of a patient undergoing transplantation for Z antitrypsin cirrhosis; inset, Z antitrypsin polymers (small arrow) formed by incubating pure plasma Z antitrypsin at 41 °C. To obtain definition of the tangled filaments the inclusion was highlighted by platinum rotatory shadowing but it shows the same ultrastructural morphology as the negatively stained (inset) polymers. *c*, Circular dichroism spectrum (far ultraviolet) of fragmented liver inclusions (dark line) and Z antitrypsin polymerized *in vitro* (unbroken line). Both show identity of structure, with retention of overall tertiary conformation and the predicted modification for loop-sheet polymerization, intermediate between native Z antitrypsin (S) and its cleaved (R) form. Denatured Z antitrypsin is also shown. Isolated liver inclusions were repeatedly washed in 50% (v:v) Tween 20 and then fragmented by glass-bead grinding in 5% Tween 20 and 10% Sarcosyl, with subsequent gel filtration to isolate the polymer fraction (average M_r 500K). Subsequent SDS-polyacrylamide gel electrophoresis showed a single band corresponding to native Z antitrypsin control.

the livers of Z homozygotes. Polyacrylamide gel electrophoresis confirmed that the Z polymers, like the liver inclusion^{10,11}, are formed by non-covalent bonding and are readily dissociable to monomers in detergents in the absence of thiol reducing agents. The critical evidence comes from the electron microscopy of the polymers (Fig. 3). Incubation of Z antitrypsin results, as predicted, in the formation of multiple bead-like polymers usually in filaments of 4–10 molecules. Examination of α_1 -antitrypsin inclusions isolated from the liver of a Z homozygote shows that they are formed of a myriad tangle of filaments, identical in structure to those seen in incubated isolated Z antitrypsin. Confirmation of this identity is provided by circular dichroism measurements in the far ultraviolet (Fig. 3c); these show superimposable spectra for both the fragmented liver inclusions and the *in vitro* polymers. Furthermore, both have the profile predictable⁴ for loop-sheet bonding, intermediate between that of intact (S) and cleaved (R) antitrypsin.

The polymerization of isolated Z antitrypsin was observed previously¹², as was the formation of intracellular aggregates¹³, but as the latter report points out, these observations do not by themselves explain the features of the disease. The occurrence of aggregation in the endoplasmic reticulum and the fact that it involves only 85% of the Z antitrypsin both follow from the mechanism of loop-sheet intermolecular bonding. This will predictably occur as the point of greatest concentration is approached in the hepatocyte just before entry to the Golgi apparatus. Furthermore, our results show that loop-sheet polymerization is an equilibrium process and there will always be some monomers as well as polymers. This accounts for the secretion of some 15% of Z antitrypsin into the plasma as functional monomers, a level that is insufficient to protect the lungs against the proteolytic damage that leads to emphysema².

The findings also explain previous observations that the number and density of the liver inclusions vary considerably between individuals, and from time to time in a single individual. As is shown here, the formation of liver aggregates will be dependent on temperature and concentration. At times of stress the formation of inclusions in the hepatocyte will be likely to overwhelm the degradative mechanisms responsible for their turnover. Antitrypsin is an acute phase protein and as such undergoes a manifold increase in production in association with temperature increases of up to 41 °C during bouts of inflammation. These stresses must contribute to the variable severity of hepatocellular involvement in children. There is now strong evidence¹⁴ to support earlier conclusions² that the liver disease of the homozygote is a direct consequence of antitrypsin accumulation and degradation in the hepatocyte. It follows that a first step in the prevention and amelioration of liver involvement in the Z homozygote infant should be the use of simple established therapies for the limitation of inflammation and pyrexia. In the longer term there is the challenging possibility of more specific interventions such as the delivery to the hepatocyte of engineered loop peptides specific to α_1 -antitrypsin. □

Received 9 March; accepted 28 April 1992.

- Laurell, C. B. & Eriksson, S. *Scand. J. Clin. Lab. Invest.* **15**, 132–140 (1963).
- Carrell, R. W. *J. Clin. Invest.* **78**, 1427–1431 (1986).
- Huber, R. & Carrell, R. W. *Biochemistry* **28**, 8951–8956 (1989).
- Carrell, R. W. & Evans, D. L. & Stein, P. E. *Nature* **353**, 576–578 (1991).
- Evans, D. L. thesis, Univ. Cambridge (1991).
- Schulze, A. J. *et al. Eur. J. Biochem.* **194**, 51–56 (1990).
- Curiel, D. T. *et al. J. Biol. Chem.* **264**, 13939–13945 (1989).
- Seyama, K. *et al. J. Biol. Chem.* **266**, 12627–12631 (1991).
- Stein, P. & Chothia, C. *J. molec. Biol.* **221**, 615–621 (1991).
- Eriksson, S. & Larsson, C. *New Engl. J. Med.* **292**, 176–180 (1975).
- Bathurst, I. C., Travis, J., George, P. M. & Carrell, R. W. *FEBS Lett.* **177**, 179–183 (1984).
- Cox, D. W., Billingsley, G. D. & Callahan, J. W. *FEBS Lett.* **205**, 255–260 (1986).
- Le, A., Ferrell, G. A., Dishon, D. S., Quyen-Quyen, A. L. & Sifers, R. N. *J. Biol. Chem.* **267**, 1072–1080 (1992).
- Carlson, J. A. *et al. J. Clin. Invest.* **83**, 1183–1190 (1989).
- Laurell, C. B., Pierce, J., Persson, U. & Thulin, E. *Eur. J. Biochem.* **57**, 107–113 (1975).

ACKNOWLEDGEMENTS. This study was supported by the Wellcome Trust and the Medical Research Council. We thank R. A. Stockley and M. Wardell for their help. BC13 peptide was synthesized by R. C. Sheppard. D.A.L. is an MRC Training Fellow and D.L.E. is an Elmore Fellow.

CORRECTIONS

Were small galaxies once the dominant cosmological population?

L. L. Cowie, A. Songaila & E. M. Hu

Nature **354**, 460–461 (1991)

IN this Letter the K correction was applied with the wrong sign. This in no way affects the conclusions or discussion, but the values of the variable M_K which appear in the table and figure must be corrected accordingly. We are grateful to W. Sutherland and R. Kron for pointing out this mistake.

Indirect chemical effects of methane on climate warming

Jos Lelieveld & Paul J. Crutzen

Nature **355**, 339–342 (1992)

IN preparing this paper, a typographical error in equation (4) has escaped our attention. The correct formulation is

$$G_{CO_2}(t) = -\frac{\partial}{\partial t} R_{CH_4}(t) \int_t^T f_{CO_2} R_{CO_2}(t' - t) dt' \quad (4)$$

Close linkage of glucokinase locus on chromosome 7p to early-onset non-insulin-dependent diabetes mellitus

Ph. Froguel, M. Vaxillaire, F. Sun, G. Velho, H. Zouali, M. O. Butel, S. Lesage, N. Vionnet, K. Clément, F. Fougerousse, Y. Tanizawa, J. Weissenbach, J. S. Beckmann, G. M. Lathrop, Ph. Passa, M. A. Permutt & D. Cohen

Nature **356**, 162–164 (1992)

THE following was omitted from the Acknowledgements section of this Letter: "This work was supported in part by a NIH grant to M.A.P. Y.T. was a recipient of a mentor based fellowship award of the American Diabetes Association."

Genetically engineered alteration in the chilling sensitivity of plants

N. Murata, O. Ishizaki-Nishizawa, S. Higashi, H. Hayashi, Y. Tasaka & I. Nishida

Nature **356**, 710–713 (1992)

THIS Letter in the 23 April issue contains errors in the figure legends, involving the definition of parts *b*, *c* and *d*. Figure 1 legend should begin "Chilling-induced damage to the photosynthetic activity of leaves of transgenic plants. *a*, Transformant with pBI-121 (control; Clontech). *b*, Transformant with pSQ. *c*, Transformant with pARA. Methods. (The same as printed.)" Figure 2 legend should begin "Chilling-induced visible damage of transgenic plants. *a*, Wild type. *b*, Transformant with pBI-121 (control). *c*, Transformant with pSQ. *d*, Transformant with pARA." □

Beyond silver staining

Andrew Wallace and Hanspeter Saluz

Proteins separated on sodium dodecyl sulphate-polyacrylamide gels can be detected in subpicogram quantities by radioactivation of silver-treated protein molecules.

THE recent resurgence of interest in protein biochemistry has placed many researchers on the track of biologically important factors which occur in nature only in small quantities. The isolation of such factors is usually a prerequisite for their identification and characterization. To follow the progress of purification in these cases, one would ideally wish to be able to use only a negligible part of the sample. Until recently, the methods commonly used to check its purity employed colorimetric detection using Coomassie blue¹ or silver staining² of the protein components separated on SDS-polyacrylamide gels³. Although not insensitive⁴, these detection procedures still required considerable amounts of precious material to be sacrificed. This shortcoming prompted us to search for a procedure that extended the sensitivity limits beyond those of silver staining.

An obvious way to increase detection sensitivity is to label biopolymers with radioactive isotopes. However, direct post-synthetic labelling of proteins by linking radioisotopes to the molecules does not ensure maintenance of their physical properties and thus might change their electrophoretic mobility. For these reasons, we sought new possibilities that would allow labelling of the proteins after their electrophoretic separation⁵.

The use of silver in detection

We chose to exploit the properties of silver, which were known from the widely-used protein staining method where insoluble deposits of silver compounds are formed around proteins separated within a gel matrix. These deposits are obtained after reduction of silver nitrate with formaldehyde², a process called development which is enhanced by the presence of protein, thus leading to a faster deposition of silver metal in the area of the protein bands compared to the rest of the gel. The concentration of silver used in this method leads to visibly stained protein bands.

In sulphur toning processes applied in photography⁶, such deposits of silver compounds are converted to the more stable silver sulphide, a procedure which is used for archiving rare photographs. Moreover, in particle- and astrophysics, ultrasensitive detection of weak photographic images was achieved by rendering silver sulphide radioactive using ³⁵S sulphur compounds in 'toning' processes⁷⁻¹¹ and exposing the now radioactively labelled photograph to an X-ray film¹².

We decided to use a similar approach for the detection of proteins in polyacrylamide gels.

Basis of the method

Our method uses the principle of protein silver staining according to Ansorge² combined with a modified sulphur toning procedure involving [³⁵S]thiourea⁵. Although several other isotopes, such as iron-55, nickel-63, silver-110m, cadmium-115m, iodine-125, promethium-147, gold-195, mercury-203, polonium-210, uranium-233, americium-241 or californium-252 would theoretically be well suited for use in a very sensitive protein detection procedure⁹, we chose the isotope ³⁵S, as it is a β -emitter that can be handled safely in the average laboratory and gives sharp signals on X-ray films.

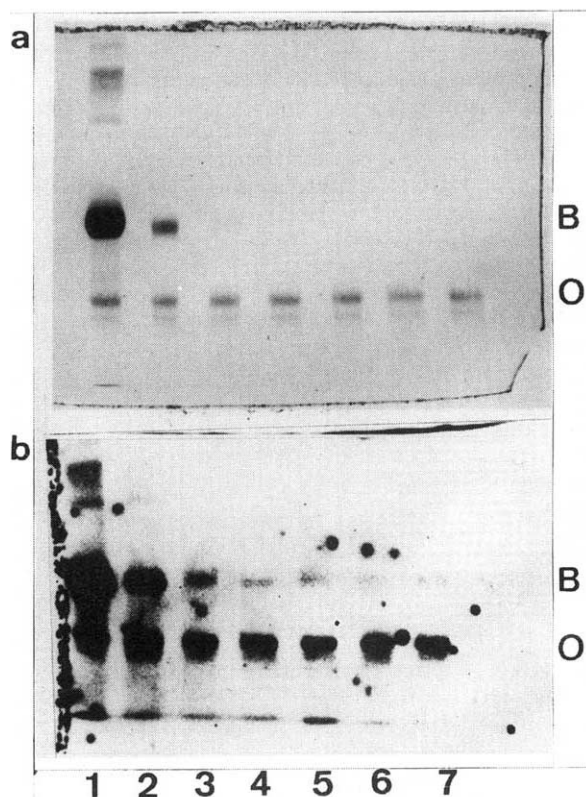
In contrast to conventional silver staining, the amount of silver we used was reduced by at least three orders of magnitude and naturally this low amount of silver did not result in visible signals. However, this drastic reduction was necessary to yield a satisfactory signal-to-background noise ratio. In order to render the deposited silver reactive with thiourea, the gels were treated with a potassium ferricyanide-potassium bromide bleach bath¹³ to form silver bromide, which was then converted to labelled silver sulphide with radioactive thiourea (obtained from Amersham, Little Chalfont, Buckinghamshire, England). The high sensitivity of the detection method is due to this radioactivation step as it introduces the parameter of time. Signal strength is, therefore, dependent on the length of exposure.

Practical application

A sensitivity comparison was made between this procedure and other commonly used staining methods, such as Coomassie brilliant blue, silver staining and a combination of both (see figure). In the figure, bovine serum albumin (BSA; MW 67,000)

was varied between 400 ng and 400 fg while ovalbumin (MW 43,000) was used as a carrier protein at a constant amount of 400 ng. In this experiment, a relatively high amount of ovalbumin carrier was mixed with the BSA sample to avoid inadvertent loss of trace quantities of protein on equipment surfaces. These protein amounts were determined according to weight and these numbers may be, if anything, an overestimate of the amounts effectively present, as salts or traces of other proteins could still contaminate the sample.

The lowest detected amount with Coomassie brilliant blue staining (panel a) was 40 ng. With the more sensitive silver staining method between 4 and 40 ng could be detected, and a combination of Coomassie brilliant blue and silver staining led to the detection of amounts down to 4 ng but with a higher signal intensity⁵ (data not shown). With the use of intensifying screens, fluorographic reagents, such as Amplify



Sensitivity comparison between different detection procedures on 10 per cent SDS-gels. (a) Coomassie blue staining and (b) radioactivation procedure. Lanes 1-7: 400 ng of ovalbumin in each track (O) and 400 ng, 40 ng, 4 ng, 400 pg, 40 pg, 4 pg, 400 fg of BSA (B).

(Amersham), and preflashing¹⁴, the radioactivation procedure (panel *b*) allowed the detection of amounts down to approximately 400 fg. (A further increase in sensitivity is expected from the utilization of the new generation of phosphorimaging equipment. Furthermore, even the film emulsions themselves can be sensitized by baking in forming gas atmosphere^{15,16}.) It can be seen from panel *b* of the figure that the response is not linear, that is, in the example shown it can be seen that small amounts of proteins show almost the same signal intensity, whereas large amounts give relatively small signals when compared with procedures involving silver staining. Similar observations were made by Owunwanne *et al.*⁸, where radioactivation of weak bands on autoradiograms resulted in a nonlinear response upon re-exposure to a fresh X-ray film.

Potential applications

An important aspect, among the many feasible applications of this novel method, is its ability to detect certain types of synthetic peptides that conventional procedures fail to stain⁵. A possible explanation for this is that the few silver grains that form around these peptides are invisible to the naked eye but are sufficient to produce detectable signals upon radioactivation. Finally, this procedure can also be used to the full potential of its high sensitivity, especially in fields where samples are unique or limited to very small amounts, such as archeology, forensic science or palaeontology. □

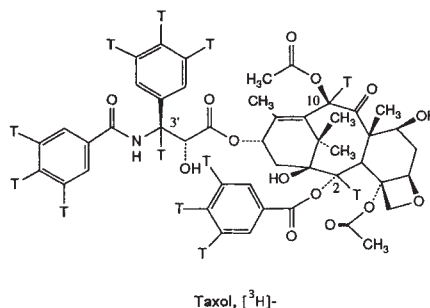
Andrew Wallace and Hanspeter Saluz are at the Istituto di Ricerche di Biologia Molecolare (IRBM), Via Pontina Km 30.600, 00040 Pomezia, Roma, Italy. For more information, fill in reader service number 100.

- Irie, S., Sezaki, M. & Kato, Y. *Analyt. Biochem.* **126**, 350–354 (1982).
- Ansorge, W. J. *Biochem. biophys. Meth.* **11**, 13–20 (1985).
- Laemmli, U. K. *Nature* **227**, 680–685 (1970).
- Merrill, C. R. *Nature* **343**, 779–780 (1990).
- Wallace, A. & Saluz, H. P. *Analyt. Biochem.* **203**, 27–34 (1992).
- Glafkides, P. in *Chimie et Physique Photographiques Vols 1 & 2* (Editions de l'Usine, Paris, 1987).
- Vachon, D., Owunwanne, A., Carroll, B. H. & O'Mara, R. E. *J. appl. fotogr. Engng* **6**, 141–143 (1980).
- Owunwanne, A., Wheaton, L. E., Carroll, B. H., O'Mara, R. E. *J. appl. fotogr. Engng* **8**, 104–106 (1982).
- Thackray, M. in *Autoradiography of Radioactive Photographic Images — Applications to Intensification, Restoration, Precision Etching, Photomechanical Reproduction and Photographic Research* 1–22 Docum. AAE/E317 (ISBN 0642996563) (International Atomic Energy Agency, Vienna, 1974).
- Askins, B. *Appl. Opt.* **15**, 2860–2865 (1976).
- Dempster, D. T. M.S. thesis, Rochester Inst. of Technol. Rochester, N.Y. (1978).
- Suzuki, S. *J. Soc. scient. Photogr. Jap.* Nov. 9–26 (1952).
- Fyson, J. R. & Levenson, G. I. P. *J. fotogr. Sci.* **28**, 194–195 (1980).
- Laskey, R. A. & Mills, A. D. J. *Biochem.* **56**, 335–341 (1975).
- Smith, A. G., Phillips, C. A. & Hahn, E. J. *J. imag. Technol.* **11**, 27–32 (1985).
- Smith, A. G., Phillips, C. A., Hahn, E. J. & Leacock, R. J. *Am. astr. Soc. Photo Bull.* **39**, 8–14 (1985).

Types of tags

Featured this week — tritium-labelled taxol, probes, conjugated monoclonal antibodies, and a system for the direct detection of beta-particles.

IN 1991, taxol was approved for clinical trials as a treatment for ovarian, breast and possible lung and colon cancers. Moravek Biochemicals now offers the antitumour agent, taxol, in a tritium-labelled form. (Reader Service No. 101). **Tritium-labelled taxol**, which has a specific activity of



Tritium-labelled taxol — see Moravek.

20–30 Ci mmol⁻¹, is available in 0.05 mCi, 0.25 mCi and 1 mCi amounts. It is shipped in an ethanol solution on dry ice. Delivery takes 1–2 days, Moravek says. The developmental work for the tritiation of taxol was supported by a grant from the National Cancer Institute.

Enzo has adapted its proprietary nonradioactive labelling technology to the **biotin-labelling of RNA transcripts** (Reader Service No. 102). Using a procedure developed at Enzo, single-stranded biotinylated RNA probes can be produced by *in vitro* transcription reactions. Biotinylated UTP (Bio-11-UTP) together with ATP, GTP and CTP in a transcription reaction catalysed by T7, T3 or SP6 RNA polymerase produce RNA transcripts biotinylated at U residues. These procedures, Enzo says, yield efficient incorporation of the biotinylated nucleotide and result in transcripts that can be used as probes or as substrates for RNA processing studies. The RNA labelling system contains reagents for the synthesis and biotin-labelling of RNA in 20 transcription reactions of 1 µg each (including controls). Two kits are available: using T3 or T7 RNA polymerase or SP6 RNA polymerase. Standard hybridization procedures can be used with the biotinylated RNA probes; the hybridized biotinylated RNA probe can then be detected with an appropriate streptavidin-biotin detection complex.

Chromosome *in situ* systems

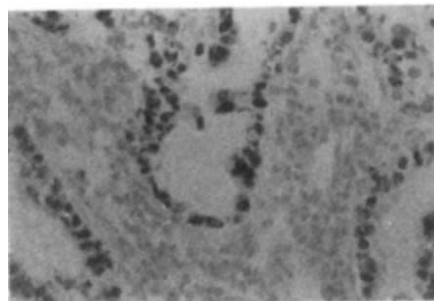
Oncor is expanding its probe families for **chromosome *in situ* hybridization** (Reader Service No. 103). The three probe families

— Unique Sequence Probes, Microdeletion Analysis Probes, Chromosome-Specific Painting Probes — are conveniently premixed with blocking DNA to suppress background binding while highlighting the specific probe target. The line-up of probes for chromosome *in situ* hybridization includes: Unique Sequence Probes (*abl* cosmid/chromosome 9 classical satellite probe, telomere 7q/chromosome 7 alpha satellite probe, telomere 4p/chromosome 4 alpha satellite probe); Human Microdeletion Analysis Probes (Wolf-Hirschhorn/chromosome 4 alpha satellite probe, Cri Du Chat probe, Miller-Dieker/chromosome 17 alpha satellite probe); and Human Chromosome-Specific Painting Probes (chromosome 7, 8, 11 and 16 specific painting probes). All of the Oncor biotin-labelled probes can be detected with either the company's fluorescence or enzyme detection kits.

Amersham has introduced three **fluorescent nucleotides for *in situ* hybridization**, which are designed to facilitate rapid, sensitive multiple hybridizations to chromosomes and nuclei (Reader Service No. 104). The range includes fluorescein (Fluorogreen), rhodamine (Fluorored) and coumarin (Fluoroblu). These fluorescent nucleotides can be incorporated into probes using nick translation or the polymerase chain reaction. Following hybridization to chromosomes or nuclei, the probe can be detected directly using fluorescence microscopy. The use of different dyes allows multiple hybridizations. Stated applications include the identification of chromosomes, whole chromosome mapping, typing karyotypes, transfer identification and single-copy identification.

Assorted monoclonals

Dako has introduced a new **monoclonal antibody to the p53 oncogene** for use in immunohistochemistry (Reader Service No. 105). Overexpression of the p53 oncogene



Lung adenocarcinoma labelled with Dako's monoclonal anti-human p53.

has been implicated as the most common genetic alteration in the development of human malignancies. Although still clinically undefined, investigations of a variety of malignancies, including neoplasms of breast, colon, ovary, lung, liver, mesenchymal, bladder and myeloid, have suggested a contributing role of p53 mutation in development of malignancy^{1,2}. Highest frequency of expression has been demonstrated in tumours of the breast, colon and ovary³. In addition, p53 mutation and overexpression has not been recognized in benign tumours or in normal tissue¹⁻³.

1. Porter, P. et al. *Am. J. Pathol.* **140**, 144-153 (1992).
2. Levine, A. J. et al. *Nature* **351**, 453-456 (1991).
3. Chank, K. et al. *J. histochem. Cytochem.* **39**, 1281-1287 (1991).

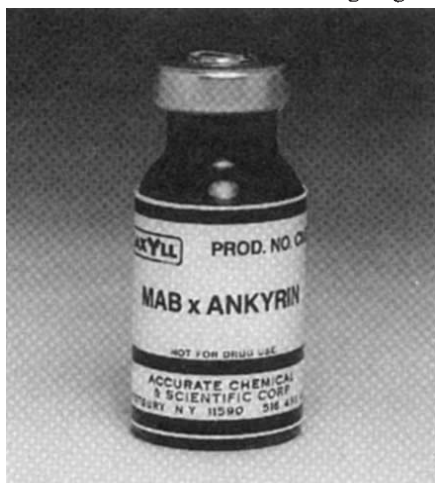
IBL Research Products has added four new products to its line of **surface and cytoskeletal monoclonal antibodies** (*Reader Service No. 106*). InnoMAb CD44, supplied either conjugated or unconjugated, is designed primarily as a cell-surface marker for detecting metastatic cancer, leukaemia, non-Hodgkins lymphomas and various lymphoid malignancies. It cross-reacts with human, rabbit, mouse and bovine cells and tissues and can be used for ELISAs, flow cytometry, immunohistostaining of paraffin-embedded tissue, immunoprecipitation and western blotting. In addition to its use in immunological and oncological studies, InnoMAb CD45RO is designed for labeling activated T-cells in drug therapy. InnoMAb Fodrin and InnoMAb Ankyrin complete the line-up. These products should be of interest to neuroscientists whose research involves aging studies.

The Finland-based company, Locus, is offering a range of **monoclonal antibodies to extracellular matrix components**, which should be of interest to researchers in cell and tumour biology as well as clinical immunopathology (*Reader Service No. 107*). The range includes monoclonal antibodies to cellular and plasma fibronectin, tenascin and laminin. Fibronectins are high molecular weight cell adhesion glycoproteins found in basement membranes, in the interstitial connective tissue matrix and in soluble form in plasma and other body fluids. Most of the fibronectin in adults is plasma fibronectin, which is produced by the liver cells. Plasma fibronectin is a soluble heterodimer. The cellular fibronectin is the insoluble tissue or extracellular matrix form of the protein or the secreted fibronectin of cultured cells. Locus has produced and characterized a monoclonal antibody that recognizes only cellular fibronectin. It is specific to the extra domain of the molecule, a part that does not exist in plasma fibronectin. Tenascin is a high molecular weight glycoprotein of connective tissue and is especially prominent in differentiating tissue. Locus has developed a monoclonal antibody to tenascin produced by human fetal fibroblasts. The com-

pany says tenascin may be used as a marker for malignancy in immunostaining, especially in the case of breast tumours. Completing the line-up is the new monoclonal antibody to laminin — the fibrous glycoprotein found in basement membranes, which is an attachment factor for epithelial cells. All three monoclonal antibodies are supplied in vials containing 50 µg of the specific affinity-purified immunoglobulin; all are also available labelled with peroxidase.

■ Anti-ankyrin agents

Monoclonal **anti-ankyrin antibody** can now be obtained from Accurate Chemical and Scientific (*Reader Service No. 108*). Ankyrin belongs to a family of closely related polypeptides associated with the plasma membrane of cells in a variety of cell types (lymphocytes, platelets, fibroblasts and endothelial cells) and various tissues (brain, kidney, muscle and many epithelial tissues). Ankyrin underlies and interacts with many different plasma membrane proteins including CD44 (H-CAM: homing cellular adhesion molecule), the voltage-dependent sodium channel, Na⁺/K⁺ ATPase and the anion exchanger protein. Accurate says, it is believed that the formation of a direct connection between ankyrin and functionally important transmembrane proteins/membrane skeleton may be one of the earliest events to occur during signal



Accurate's anti-ankyrin monoclonal.

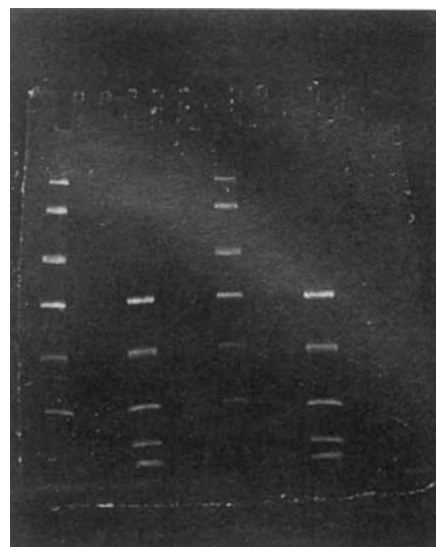
transduction and cell activation. The anti-ankyrin monoclonal antibody is suitable for use in procedures involving western blotting, ELISAs, immunoprecipitation, immunocytochemistry and flow cytometry. It can also be conjugated to numerous labels, including enzymes, fluorochromes, biotin or agarose.

Ankyrin consists of a single 215-kD peptide that is known to be involved in the attachment of spectrin- or fodrin-containing membrane skeletons to the plasma membrane. Zymed's new **mouse monoclonal antibody to ankyrin** was raised against purified human erythroid cell ankyrin and

purified from hybridoma supernatant (*Reader Service No. 109*). It reacts specifically with the 215-kD intracellular ankyrin in human, mouse and bovine tissues and can be used in procedures involving immunoprecipitation, western blots, ELISAs and immunofluorescence.

■ Go west!

Intended for western blotting, the new **fluorescent protein standards** from Integrated Separation Systems are visualized



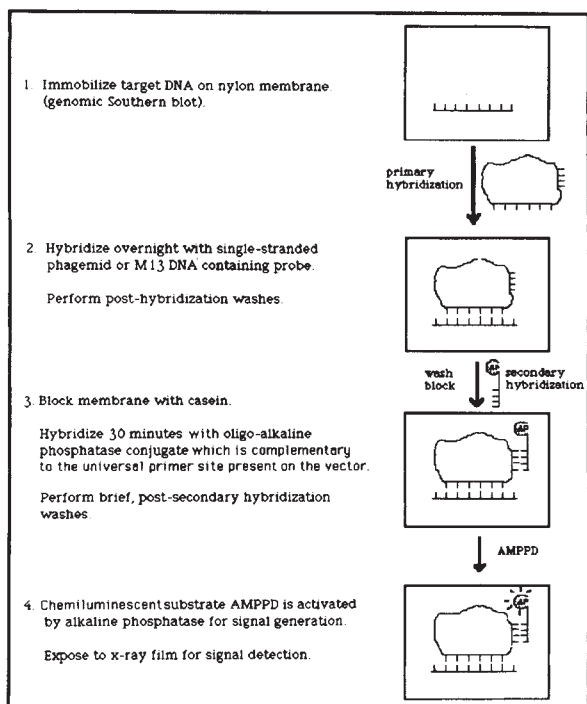
Fluorescent protein standards from ISS.

with UV light after transfer to membrane or during polyacrylamide gel electrophoresis (*Reader Service No. 110*). ISS says fluorescent standards run true to their molecular weight and can be used for accurate calculation of R_F values. Overstaining with total protein stains, such as Coomassie blue or silver, results in sharp bands without the appearance of contaminating bands. Mid-range standards have six bands between 95 and 14 kD; low-range standards have five bands between 29 and 3.5 kD.

Western-Light is the new **nonisotopic western blotting system** from Tropix that is intended to provide increased sensitivity, flexibility and control over film exposure (*Reader Service No. 111*). According to Tropix, results are obtained in minutes with X-ray or instant film. And, fast film exposures can be obtained on nitrocellulose, PVDF or nylon membranes. Western-Light is available as a complete protein labelling and chemiluminescent detection system. This kit, and a host of other new products for nonradioactive analysis of nucleic acids and proteins, are featured in the company's 1992/1993 product catalogue.

■ Probe detection kits

Gene-Lite is the new kit from Bio-Rad that is designed for applications involving high sensitivity, **single-copy gene detection** (*Reader Service No. 112*). The procedure is



Gene-Lite — Bio-Rad's chemiluminescent nucleic acid detection kit

a two-step hybridization process. First, an unlabelled single-stranded primary probe is detected by an alkaline phosphatase-conjugated oligonucleotide secondary probe. Second, the conjugate activates the chemiluminescent substrate, which generates autoradiograms with low background and high signal after overnight exposure with as little as 2 µg of genomic DNA, Bio-Rad says. The two-step procedure is designed to eliminate the problem of inadequately labelled probes and the high backgrounds associated with avidin or anti-digoxigenin alkaline phosphatase conjugates.

British Bio-technology has extended its range of *in situ* hybridization products to include a **detection module for biotinylated probes** (Reader Service No. 113). BBT's kit offers a method of detecting biotin-labelled oligonucleotides that have been annealed to target mRNA by *in situ* hybridization. These probes are then detected using an avidin/alkaline phosphatase conjugate, which, after addition of the revealing agents, produces an intense blue/black precipitate at the site of hybridization. BBT says the biotin labelling and detection system offers several advantages over conventional methods: the probes are stable and non-radioactive, and the system offers good cellular resolution and fast results. The entire procedure is designed to be completed within two days, compared to weeks when using emulsion autoradiography with radioactive probes.

BioGenex Laboratories says its Super Sensitive **DNA probe detection system**, NATURE • VOL 357 • 18 JUNE 1992

which incorporates the company's biotin-streptavidin amplified technology, provides up to 15 times the sensitivity of conventional avidin-streptavidin technology-based products (Reader Service No. 114). This nonradioactive system for *in situ* hybridization is capable of detecting as few as one to two copies of target DNA. After hybridization to target DNA *in situ*, the biotinylated nucleic acid probe is visualized using an amplified immunoassay. Amplification is achieved by the sequential addition of anti-biotin IgG, biotinylated anti-IgG, and an alkaline phosphatase-conjugated streptavidin label. Each step provides an additional layer of amplification, with the increased sensitivity permitting up to 15-fold dilution of commercially available probes. Each detection kit includes all the ready-to-use reagents necessary to

develop 50 *in situ* hybridization assays on formalin-fixed, paraffin-embedded tissue sections. BioGenex also supplies a High Performance DNA probe detection system. The company says that this is a simplified system for the rapid detection of biotinylated probes, which provides up to five times the sensitivity of comparable avidin-biotin products.

■ Means of measurement

Eliminate the need for autoradiography, densitometric analysis or scintillation counting, says Ambis, with the new Ambis 4000 — an instrument for the **direct detection of beta particles** emitted from ^{14}C , ^{35}S , ^{32}P , ^{125}I and a variety of other isotopes (Reader Service No. 115). The Ambis 4000 uses a direct counting method and provides the researcher with true nuclear counting statistics. It features an optional sample changer that allows the user to automate the imaging process still further. Another key feature is the integrated scanner keypad that allows the user to operate the instrument independ-



Ambis 4000 for the direct detection of beta particles.

ently of the computer system. Remote data analysis workstations (IBM PC/AT compatible) are offered to encourage the user to analyse sample data outside of the laboratory, eliminating the need for a central analysis workstation. The modular nature of the Ambis 4000 system allows the user to customize the instrument to suit specific needs.

QC-Scan, the new compact thin layer chromatography (TLC) scanner from Bioscan, is designed for the rapid **quantitative analysis of all gamma- and beta-emitting isotopes on TLC plates**, iTLC strips and paper chromatograms (Reader Service No. 116). For most radiopurity analyses, results can be obtained in less than a minute, Bioscan says. The whole system — microprocessor-controlled scanning protocols and data reduction, keyboard and



QC-Scan provides radiopurity analysis for thin layer chromatography plates.

built-in graphics printer — requires only one-square foot of bench space and fits conveniently on the laboratory bench or under a radiosynthesis hood. Three versions of the instrument are available. For radiopharmaceutical quality control, the QC-Scan gamma version detects gamma-emitting isotopes with energies up to 400 keV, such as $^{99\text{m}}\text{Tc}$, ^{131}I and ^{111}In . For radiochemical purity analysis, the two QC-Scan beta versions detect low-energy gamma- and beta-emitting isotopes, including ^3H , ^{14}C , ^{32}P and ^{125}I . A paper-strip version has two detectors and can scan strips up to 100 cm in length.

MGM Industries offers a choice of two **luminometers** — one manual, one automated — which are optimized for high-sensitivity bio- and chemiluminescence applications (Reader Service No. 117). With Optocomp I, a manual, photon-counting luminometer, automatic injection, count timing, data acquisition and printing are initiated by closing the lid. The Optocomp II, however, is a cassette-based automated version that is designed to process up to 250 samples in 12 × 75 test tubes, or up to 400 microtubes, without the need for operator intervention. Operation of both models,

ADVERTISEMENTS

Highest Quality Oligonucleotides
 Synthesised, purified and shipped in 3 days
£2.75 per base
 (All inclusive)
 0.2 µM scale (approx. 500 µg)

- Antisense (completely or partially sulphurised)
- Extensive range of labelling and modifications
- Full design and analysis facilities
- Gene cassettes, bulk order and contract discounts

Molecular Medicine Unit
 King's College School of Medicine & Dentistry
 Tel. 071 - 326 3126 FAX 071 - 733 3877

Reader Service No 56

\$3.50/Base
Custom DNA Synthesis
 No Set-up Fee,
 Shipped in 24 hours.
1-800-788-4362
1-800-786-4362 Fax

KEYSTONE
 Laboratories, Inc.

Reader Service No.33

Synthetic Oligonucleotides
 Rapid Quality Worldwide Service
£2.30 per base, no setup!
 (0.2 µmole - limited offer)
 option of H.P.L.C. purification - £30
 biotinylation, fluorescent labelling
 phosphothioates, many other modifications

OSWEL DNA SERVICE
 Department of Chemistry, University of Edinburgh
 Kings Buildings, West Mains Road,
 EDINBURGH EH9 3JJ
 Tel: 031 650 4792 Fax: 031 662 4054

Reader Service No.9

HPLC, TLC, GC

RADIOACTIVITY + CHROMATOGRAPHY

raytest

raytest • Isotopenmeßgeräte GmbH
 D-7541 Straubenhardt 1 • Benzstraße 4
 Tel. 07082 / 6 04 07 • Fax 07082 / 2 08 13

...WE LISTEN TO OUR CUSTOMERS

Reader Service No.29

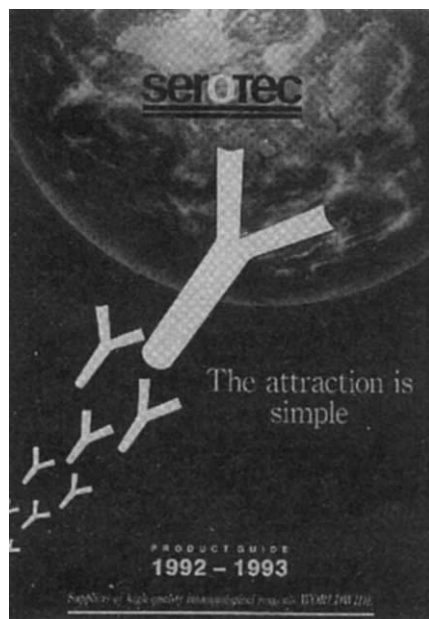


Optocomp I manual luminometer.

MGM says, can be learned in 5 mins. Software capabilities include kinetics, raw data and a variety of curve-fitting routines. Up to 30 protocols can be set up and an analysis can be started by pushing only three buttons. Incubation software is also available for the Optocomp II for long-term kinetic studies or for luminescence reagents needing incubation after injection of the initiating reagent. Both models feature up to three injectors, each with individually controlled timing. A delay can also be set between the last injection and the start of measurement. A built-in printer/plotter reports the data and plots the standard curves. External laboratory computers can be interfaced through the RS-232 port on the luminometer. This bi-directional data exchange also allows the luminometer to be controlled from an external source.

■ Off the shelf

Pick up a copy of the 1992 Serotec product guide, which offers a wide selection of **immunological reagents** (Reader Service No. 118). Some of the new product highlights include an extensive range of monoclonal antibodies to human leukocytes, both conjugated (to fluorescein



Serotec's guide to new products.

isothiocyanate, RPE and biotin) and unconjugated; cytokine antibodies and recombinant natural cytokines; monoclonal antibodies to mouse leukocyte markers; monoclonal antibodies to guinea pig antigens; and a freeze drying service.

The new **catalogue and handbook of immunogold silver staining** from BioCell contains information and advice on how to choose a gold conjugate and how to obtain the best results with immunogold reagents (Reader Service No. 119). Researchers who are new to these techniques should find the introduction to immunogold staining a useful reference guide; more experienced users should find the later chapters useful as they provide a more detailed account of the use of gold labels in electron microscopy, light microscopy and immunoblotting applications. BioCell's product line includes reagents for *in situ* hybridization and ultra-small (1-nm) gold conjugates. Other sections cover methods of staining total proteins in western and dot blots with gold colloids, as well as the use of silver enhancing for microscopy and blotting. A list of relevant literature on the subject is also offered.

■ Reagent round-up

ICN Flow introduces the **protein labelling reagent**, Trans³⁵S Label, which is designed to provide researchers with an economical source of [³⁵S]methionine (Reader Service No. 120). Trans³⁵S Label, which contains both [³⁵S]methionine (70 per cent) and [³⁵S]cysteine (15 per cent), is intended for use in metabolic studies. ICN says that, to date, published studies report successful work with mammalian, yeast and insect cells. High-activity protein labelling is assured with Trans³⁵S Label as the [³⁵S]methionine component is supplied at a specific activity of greater than 1,000 Ci mmol⁻¹ at a concentration of 10 mCi ml⁻¹, ICN says. Each batch of Trans³⁵S Label is biologically tested to guarantee performance and is supplied in 50 mM L-lysine stabilizing buffer.

RapiDiff, the new multipurpose **stain for blood and bone marrow cells** from Cytocolor, is now being distributed in the United Kingdom by Advanced Laboratory Techniques (Reader Service No. 121). With the RapiDiff multipurpose stain, blood and bone marrow cells stain brightly and permanently in about a minute, the company says. Nuclear and cytoplasmic detail can be seen. The stain also provides good coloration of fine needle aspirates. The complete kit contains two solution bottles and lyophilized buffer — sufficient reagents to stain 500 slides. □

These notes are compiled by Diane Gershon from information provided by the manufacturers. To obtain further details, use the reader service card bound inside the journal.